UNITED STATES AIR FORCE

SUMMER RESEARCH PROGRAM -- 1996

SUMMER FACULTY RESEARCH PROGRAM FINAL REPORTS

VOLUME 2B

ARMSTRONG LABORATORY

RESEARCH & DEVELOPMENT LABORATORIES

5800 Uplander Way

Culver City, CA 90230-6608

Program Director, RDL
Gary Moore

Program Manager, AFOSR
Major Linda Steel-Goodwin

Program Manager, RDL
Scott Licoscos

Program Administrator, RDL
Johnetta Thompson

Program Administrator, RDL
Rebecca Kelly

Submitted to:

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH

Bolling Air Force Base

Washington, D.C.

December 1996

REPORT DOCUMENTATION PAGE

AFRL-SR-BL-TR-00-

0728

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REF
	December, 1996	

4. TITLE AND SUBTITLE
1996 Summer Research Program (SRP), Summer Faculty Research Program (SFRP), Final Reports, Volume 2B, Armstrong Laboratory

5. FUNDING NUMBERS
F49620-93-C-0063

6. AUTHOR(S)
Gary Moore

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Research & Development Laboratories (RDL)
5800 Uplander Way
Culver City, CA 90230-6608

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Office of Scientific Research (AFOSR)
801 N. Randolph St.
Arlington, VA 22203-1977

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION AVAILABILITY STATEMENT
Approved for Public Release

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*
The United States Air Force Summer Research Program (USAF-SRP) is designed to introduce university, college, and technical institute faculty members, graduate students, and high school students to Air Force research. This is accomplished by the faculty members (Summer Faculty Research Program, (SFRP)), graduate students (Graduate Student Research Program (GSRP)), and high school students (High School Apprenticeship Program (HSAP)) being selected on a nationally advertised competitive basis during the summer intersession period to perform research at Air Force Research Laboratory (AFRL) Technical Directorates, Air Force Air Logistics Centers (ALC), and other AF Laboratories. This volume consists of a program overview, program management statistics, and the final technical reports from the SFRP participants at the Armstrong Laboratory.

14. SUBJECT TERMS
Air Force Research, Air Force, Engineering, Laboratories, Reports, Summer, Universities, Faculty, Graduate Student, High School Student

15. NUMBER OF PAGES

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL

PREFACE

Reports in this volume are numbered consecutively beginning with number 1. Each report is paginated with the report number followed by consecutive page numbers, e.g., 1-1, 1-2, 1-3; 2-1, 2-2, 2-3.

Due to its length, Volume 2 is bound in two parts, 2A and 2B. Volume 2A contains #1-22. Volume 2B contains reports #23-43. The Table of Contents for Volume 2 is included in both parts.

This document is one of a set of 16 volumes describing the 1996 AFOSR Summer Research Program. The following volumes comprise the set:

VOLUME	TITLE
1	Program Management Report
	Summer Faculty Research Program (SFRP) Reports
2A & 2B	Armstrong Laboratory
3A & 3B	Phillips Laboratory
4	Rome Laboratory
5A, 5B & 5C	Wright Laboratory
6	Arnold Engineering Development Center, Wilford Hall Medical Center and Air Logistics Centers
	Graduate Student Research Program (GSRP) Reports
7A & 7B	Armstrong Laboratory
8	Phillips Laboratory
9	Rome Laboratory
10A & 10B	Wright Laboratory
11	Arnold Engineering Development Center, United States Air Force Academy, Wilford Hall Medical Center, and Wright Patterson Medical Center
	High School Apprenticeship Program (HSAP) Reports
12A & 12B	Armstrong Laboratory
13	Phillips Laboratory
14	Rome Laboratory
15A&15B	Wright Laboratory
16	Arnold Engineering Development Center

SRP Final Report Table of Contents

SRP Final Report Table of Contents

SRP Final Report Table of Contents

SRP Final Report Table of Contents

SRP Final Report Table of Contents

SRP Final Report Table of Contents

SRP Final Report Table of Contents

SRP Final Report Table of Contents

SRP Final Report Table of Contents

SRP Final Report Table of Contents

SRP Final Report Table of Contents

SRP Final Report Table of Contents

SRP Final Report Table of Contents

SRP Final Report Table of Contents

SRP Final Report Table of Contents

SRP Final Report Table of Contents

SRP Final Report Table of Contents

INTRODUCTION

The Summer Research Program (SRP), sponsored by the Air Force Office of Scientific Research (AFOSR), offers paid opportunities for university faculty, graduate students, and high school students to conduct research in U.S. Air Force research laboratories nationwide during the summer.

Introduced by AFOSR in 1978, this innovative program is based on the concept of teaming academic researchers with Air Force scientists in the same disciplines using laboratory facilities and equipment not often available at associates' institutions.

The Summer Faculty Research Program (SFRP) is open annually to approximately 150 faculty members with at least two years of teaching and/or research experience in accredited U.S. colleges, universities, or technical institutions. SFRP associates must be either U.S. citizens or permanent residents.

The Graduate Student Research Program (GSRP) is open annually to approximately 100 graduate students holding a bachelor's or a master's degree; GSRP associates must be U.S. citizens enrolled full time at an accredited institution.

The High School Apprentice Program (HSAP) annually selects about 125 high school students located within a twenty mile commuting distance of participating Air Force laboratories.

AFOSR also offers its research associates an opportunity, under the Summer Research Extension Program (SREP), to continue their AFOSR-sponsored research at their home institutions through the award of research grants. In 1994 the maximum amount of each grant was increased from $20,000 to $25,000, and the number of AFOSR-sponsored grants decreased from 75 to 60. A separate annual report is compiled on the SREP.

The numbers of projected summer research participants in each of the three categories and SREP "grants" are usually increased through direct sponsorship by participating laboratories.

AFOSR's SRP has well served its objectives of building critical links between Air Force research laboratories and the academic community, opening avenues of communications and forging new research relationships between Air Force and academic technical experts in areas of national interest, and strengthening the nation's efforts to sustain careers in science and engineering. The success of the SRP can be gauged from its growth from inception (see Table 1) and from the favorable responses the 1996 participants expressed in end-of-tour SRP evaluations (Appendix B).

AFOSR contracts for administration of the SRP by civilian contractors. The contract was first awarded to Research & Development Laboratories (RDL) in September 1990. After

completion of the 1990 contract, RDL (in 1993) won the recompetition for the basic year and four 1-year options.

2. PARTICIPATION IN THE SUMMER RESEARCH PROGRAM

The SRP began with faculty associates in 1979; graduate students were added in 1982 and high school students in 1986. The following table shows the number of associates in the program each year.

YEAR	SRP Participation, by Year			TOTAL
	SFRP	GSRP	HSAP	
1979	70			70
1980	87			87
1981	87			87
1982	91	17		108
1983	101	53		154
1984	152	84		236
1985	154	92		246
1986	158	100	42	300
1987	159	101	73	333
1988	153	107	101	361
1989	168	102	103	373
1990	165	121	132	418
1991	170	142	132	444
1992	185	121	159	464
1993	187	117	136	440
1994	192	117	133	442
1995	190	115	137	442
1996	188	109	138	435

Beginning in 1993, due to budget cuts, some of the laboratories weren't able to afford to fund as many associates as in previous years. Since then, the number of funded positions has remained fairly constant at a slightly lower level.

3. RECRUITING AND SELECTION

The SRP is conducted on a nationally advertised and competitive-selection basis. The advertising for faculty and graduate students consisted primarily of the mailing of 8,000 52-page SRP brochures to chairpersons of departments relevant to AFOSR research and to administrators of grants in accredited universities, colleges, and technical institutions. Historically Black Colleges and Universities (HBCUs) and Minority Institutions (MIs) were included. Brochures also went to all participating USAF laboratories, the previous year's participants, and numerous individual requesters (over 1000 annually).

RDL placed advertisements in the following publications: *Black Issues in Higher Education, Winds of Change,* and *IEEE Spectrum.* Because no participants list either *Physics Today* or *Chemical & Engineering News* as being their source of learning about the program for the past several years, advertisements in these magazines were dropped, and the funds were used to cover increases in brochure printing costs.

High school applicants can participate only in laboratories located no more than 20 miles from their residence. Tailored brochures on the HSAP were sent to the head counselors of 180 high schools in the vicinity of participating laboratories, with instructions for publicizing the program in their schools. High school students selected to serve at Wright Laboratory's Armament Directorate (Eglin Air Force Base, Florida) serve eleven weeks as opposed to the eight weeks normally worked by high school students at all other participating laboratories.

Each SFRP or GSRP applicant is given a first, second, and third choice of laboratory. High school students who have more than one laboratory or directorate near their homes are also given first, second, and third choices.

Laboratories make their selections and prioritize their nominees. AFOSR then determines the number to be funded at each laboratory and approves laboratories' selections.

Subsequently, laboratories use their own funds to sponsor additional candidates. Some selectees do not accept the appointment, so alternate candidates are chosen. This multi-step selection procedure results in some candidates being notified of their acceptance after scheduled deadlines. The total applicants and participants for 1996 are shown in this table.

1996 Applicants and Participants			
PARTICIPANT CATEGORY	TOTAL APPLICANTS	SELECTEES	DECLINING SELECTEES
SFRP	572	188	39
(HBCU/MI)	(119)	(27)	(5)
GSRP	235	109	7
(HBCU/MI)	(18)	(7)	(1)
HSAP	474	138	8
TOTAL	1281	435	54

4. SITE VISITS

During June and July of 1996, representatives of both AFOSR/NI and RDL visited each participating laboratory to provide briefings, answer questions, and resolve problems for both laboratory personnel and participants. The objective was to ensure that the SRP would be as constructive as possible for all participants. Both SRP participants and RDL representatives found these visits beneficial. At many of the laboratories, this was the only opportunity for all participants to meet at one time to share their experiences and exchange ideas.

5. HISTORICALLY BLACK COLLEGES AND UNIVERSITIES AND MINORITY INSTITUTIONS (HBCU/MIs)

Before 1993, an RDL program representative visited from seven to ten different HBCU/MIs annually to promote interest in the SRP among the faculty and graduate students. These efforts were marginally effective, yielding a doubling of HBCI/MI applicants. In an effort to achieve AFOSR's goal of 10% of all applicants and selectees being HBCU/MI qualified, the RDL team decided to try other avenues of approach to increase the number of qualified applicants. Through the combined efforts of the AFOSR Program Office at Bolling AFB and RDL, two very active minority groups were found, HACU (Hispanic American Colleges and Universities) and AISES (American Indian Science and Engineering Society). RDL is in communication with representatives of each of these organizations on a monthly basis to keep up with the their activities and special events. Both organizations have widely-distributed magazines/quarterlies in which RDL placed ads.

Since 1994 the number of both SFRP and GSRP HBCU/MI applicants and participants has increased ten-fold, from about two dozen SFRP applicants and a half dozen selectees to over 100 applicants and two dozen selectees, and a half-dozen GSRP applicants and two or three selectees to 18 applicants and 7 or 8 selectees. Since 1993, the SFRP had a two-fold applicant

increase and a two-fold selectee increase. Since 1993, the GSRP had a three-fold applicant increase and a three to four-fold increase in selectees.

In addition to RDL's special recruiting efforts, AFOSR attempts each year to obtain additional funding or use leftover funding from cancellations the past year to fund HBCU/MI associates. This year, 5 HBCU/MI SFRPs declined after they were selected (and there was no one qualified to replace them with). The following table records HBCU/MI participation in this program.

SRP HBCU/MI Participation, By Year				
YEAR	SFRP		GSRP	
	Applicants	Participants	Applicants	Participants
1985	76	23	15	11
1986	70	18	20	10
1987	82	32	32	10
1988	53	17	23	14
1989	39	15	13	4
1990	43	14	17	3
1991	42	13	8	5
1992	70	13	9	5
1993	60	13	6	2
1994	90	16	11	6
1995	90	21	20	8
1996	119	27	18	7

6. SRP FUNDING SOURCES

Funding sources for the 1996 SRP were the AFOSR-provided slots for the basic contract and laboratory funds. Funding sources by category for the 1996 SRP selected participants are shown here.

1996 SRP FUNDING CATEGORY	SFRP	GSRP	HSAP
AFOSR Basic Allocation Funds	141	85	123
USAF Laboratory Funds	37	19	15
HBCU/MI By AFOSR (Using Procured Addn'l Funds)	10	5	0
TOTAL	188	109	138

SFRP - 150 were selected, but nine canceled too late to be replaced.
GSRP - 90 were selected, but five canceled too late to be replaced (10 allocations for the ALCs were withheld by AFOSR.)
HSAP - 125 were selected, but two canceled too late to be replaced.

7. COMPENSATION FOR PARTICIPANTS

Compensation for SRP participants, per five-day work week, is shown in this table.

1996 SRP Associate Compensation

PARTICIPANT CATEGORY	1991	1992	1993	1994	1995	1996
Faculty Members	$690	$718	$740	$740	$740	$770
Graduate Student (Master's Degree)	$425	$442	$455	$455	$455	$470
Graduate Student (Bachelor's Degree)	$365	$380	$391	$391	$391	$400
High School Student (First Year)	$200	$200	$200	$200	$200	$200
High School Student (Subsequent Years)	$240	$240	$240	$240	$240	$240

The program also offered associates whose homes were more than 50 miles from the laboratory an expense allowance (seven days per week) of $50/day for faculty and $40/day for graduate students. Transportation to the laboratory at the beginning of their tour and back to their home destinations at the end was also reimbursed for these participants. Of the combined SFRP and

GSRP associates, 65 % (194 out of 297) claimed travel reimbursements at an average round-trip cost of $780.

Faculty members were encouraged to visit their laboratories before their summer tour began. All costs of these orientation visits were reimbursed. Forty-five percent (85 out of 188) of faculty associates took orientation trips at an average cost of $444. By contrast, in 1993, 58 % of SFRP associates took orientation visits at an average cost of $685; that was the highest percentage of associates opting to take an orientation trip since RDL has administered the SRP, and the highest average cost of an orientation trip. These 1993 numbers are included to show the fluctuation which can occur in these numbers for planning purposes.

Program participants submitted biweekly vouchers countersigned by their laboratory research focal point, and RDL issued paychecks so as to arrive in associates' hands two weeks later.

In 1996, RDL implemented direct deposit as a payment option for SFRP and GSRP associates. There were some growing pains. Of the 128 associates who opted for direct deposit, 17 did not check to ensure that their financial institutions could support direct deposit (and they couldn't), and eight associates never did provide RDL with their banks' ABA number (direct deposit bank routing number), so only 103 associates actually participated in the direct deposit program. The remaining associates received their stipend and expense payments via checks sent in the US mail.

HSAP program participants were considered actual RDL employees, and their respective state and federal income tax and Social Security were withheld from their paychecks. By the nature of their independent research, SFRP and GSRP program participants were considered to be consultants or independent contractors. As such, SFRP and GSRP associates were responsible for their own income taxes, Social Security, and insurance.

8. CONTENTS OF THE 1996 REPORT

The complete set of reports for the 1996 SRP includes this program management report (Volume 1) augmented by fifteen volumes of final research reports by the 1996 associates, as indicated below:

1996 SRP Final Report Volume Assignments

LABORATORY	SFRP	GSRP	HSAP
Armstrong	2	7	12
Phillips	3	8	13
Rome	4	9	14
Wright	5A, 5B	10	15
AEDC, ALCs, WHMC	6	11	16

APPENDIX A – PROGRAM STATISTICAL SUMMARY

A. Colleges/Universities Represented

Selected SFRP associates represented 169 different colleges, universities, and institutions, GSRP associates represented 95 different colleges, universities, and institutions.

B. States Represented

SFRP -Applicants came from 47 states plus Washington D.C. and Puerto Rico. Selectees represent 44 states plus Puerto Rico.

GSRP - Applicants came from 44 states and Puerto Rico. Selectees represent 32 states.

HSAP - Applicants came from thirteen states. Selectees represent nine states.

Total Number of Participants	
SFRP	188
GSRP	109
HSAP	138
TOTAL	435

Degrees Represented			
	SFRP	GSRP	TOTAL
Doctoral	184	1	185
Master's	4	48	52
Bachelor's	0	60	60
TOTAL	188	109	297

SFRP Academic Titles	
Assistant Professor	79
Associate Professor	59
Professor	42
Instructor	3
Chairman	0
Visiting Professor	1
Visiting Assoc. Prof.	0
Research Associate	4
TOTAL	188

Source of Learning About the SRP		
Category	Applicants	Selectees
Applied/participated in prior years	28%	34%
Colleague familiar with SRP	19%	16%
Brochure mailed to institution	23%	17%
Contact with Air Force laboratory	17%	23%
IEEE Spectrum	2%	1%
BIIHE	1%	1%
Other source	10%	8%
TOTAL	100%	100%

APPENDIX B -- SRP EVALUATION RESPONSES

1. OVERVIEW

Evaluations were completed and returned to RDL by four groups at the completion of the SRP. The number of respondents in each group is shown below.

Table B-1. Total SRP Evaluations Received

Evaluation Group	Responses
SFRP & GSRPs	275
HSAPs	113
USAF Laboratory Focal Points	84
USAF Laboratory HSAP Mentors	6

All groups indicate unanimous enthusiasm for the SRP experience.

The summarized recommendations for program improvement from both associates and laboratory personnel are listed below:

A. Better preparation on the labs' part prior to associates' arrival (i.e., office space, computer assets, clearly defined scope of work).

B. Faculty Associates suggest higher stipends for SFRP associates.

C. Both HSAP Air Force laboratory mentors and associates would like the summer tour extended from the current 8 weeks to either 10 or 11 weeks; the groups state it takes 4-6 weeks just to get high school students up-to-speed on what's going on at laboratory. (Note: this same argument was used to raise the faculty and graduate student participation time a few years ago.)

2. 1996 USAF LABORATORY FOCAL POINT (LFP) EVALUATION RESPONSES

The summarized results listed below are from the 84 LFP evaluations received.

1. LFP evaluations received and associate preferences:

Table B-2. Air Force LFP Evaluation Responses (By Type)

		How Many Associates Would You Prefer To Get ?								(% Response)			
		SFRP				GSRP (w/Univ Professor)				GSRP (w/o Univ Professor)			
Lab	Evals Recv'd	0	1	2	3+	0	1	2	3+	0	1	2	3+
AEDC	0	-	-	-	-	-	-	-	-	-	-	-	-
WHMC	0	-	-	-	-	-	-	-	-	-	-	-	-
AL	7	28	28	28	14	54	14	28	0	86	0	14	0
FJSRL	1	0	100	0	0	100	0	0	0	0	100	0	0
PL	25	40	40	16	4	88	12	0	0	84	12	4	0
RL	5	60	40	0	0	80	10	0	0	100	0	0	0
WL	46	30	43	20	6	78	17	4	0	93	4	2	0
Total	84	32%	50%	13%	5%	80%	11%	6%	0%	73%	23%	4%	0%

LFP Evaluation Summary. The summarized responses, by laboratory, are listed on the following page. LFPs were asked to rate the following questions on a scale from 1 (below average) to 5 (above average).

2. LFPs involved in SRP associate application evaluation process:
   a. Time available for evaluation of applications:
   b. Adequacy of applications for selection process:
3. Value of orientation trips:
4. Length of research tour:
5 a. Benefits of associate's work to laboratory:
   b. Benefits of associate's work to Air Force:
6. a. Enhancement of research qualifications for LFP and staff:
   b. Enhancement of research qualifications for SFRP associate:
   c. Enhancement of research qualifications for GSRP associate:
7. a. Enhancement of knowledge for LFP and staff:
   b. Enhancement of knowledge for SFRP associate:
   c. Enhancement of knowledge for GSRP associate:
8. Value of Air Force and university links:
9. Potential for future collaboration:
10. a. Your working relationship with SFRP:
    b. Your working relationship with GSRP:
11. Expenditure of your time worthwhile:

(Continued on next page)

12. Quality of program literature for associate:
13. a. Quality of RDL's communications with you:
    b. Quality of RDL's communications with associates:
14. Overall assessment of SRP:

Table B-3. Laboratory Focal Point Reponses to above questions

	AEDC	*AL*	*FJSRL*	*PL*	*RL*	*WHMC*	*WL*
# Evals Recv'd	0	7	1	14	5	0	46
Question #							
2	-	86 %	0 %	88 %	80 %	-	85 %
2a	-	4.3	n/a	3.8	4.0	-	3.6
2b	-	4.0	n/a	3.9	4.5	-	4.1
3	-	4.5	n/a	4.3	4.3	-	3.7
4	-	4.1	4.0	4.1	4.2	-	3.9
5a	-	4.3	5.0	4.3	4.6	-	4.4
5b	-	4.5	n/a	4.2	4.6	-	4.3
6a	-	4.5	5.0	4.0	4.4	-	4.3
6b	-	4.3	n/a	4.1	5.0	-	4.4
6c	-	3.7	5.0	3.5	5.0	-	4.3
7a	-	4.7	5.0	4.0	4.4	-	4.3
7b	-	4.3	n/a	4.2	5.0	-	4.4
7c	-	4.0	5.0	3.9	5.0	-	4.3
8	-	4.6	4.0	4.5	4.6	-	4.3
9	-	4.9	5.0	4.4	4.8	-	4.2
10a	-	5.0	n/a	4.6	4.6	-	4.6
10b	-	4.7	5.0	3.9	5.0	-	4.4
11	-	4.6	5.0	4.4	4.8	-	4.4
12	-	4.0	4.0	4.0	4.2	-	3.8
13a	-	3.2	4.0	3.5	3.8	-	3.4
13b	-	3.4	4.0	3.6	4.5	-	3.6
14	-	4.4	5.0	4.4	4.8	-	4.4

3. 1996 SFRP & GSRP EVALUATION RESPONSES

The summarized results listed below are from the 257 SFRP/GSRP evaluations received.

Associates were asked to rate the following questions on a scale from 1 (below average) to 5 (above average) - by Air Force base results and over-all results of the 1996 evaluations are listed after the questions.

1. The match between the laboratories research and your field:
2. Your working relationship with your LFP:
3. Enhancement of your academic qualifications:
4. Enhancement of your research qualifications:
5. Lab readiness for you: LFP, task, plan:
6. Lab readiness for you: equipment, supplies, facilities:
7. Lab resources:
8. Lab research and administrative support:
9. Adequacy of brochure and associate handbook:
10. RDL communications with you:
11. Overall payment procedures:
12. Overall assessment of the SRP:
13.
    a. Would you apply again?
    b. Will you continue this or related research?
14. Was length of your tour satisfactory?
15. Percentage of associates who experienced difficulties in finding housing:
16. Where did you stay during your SRP tour?
    a. At Home:
    b. With Friend:
    c. On Local Economy:
    d. Base Quarters:
17. Value of orientation visit:
    a. Essential:
    b. Convenient:
    c. Not Worth Cost:
    d. Not Used:

SFRP and GSRP associate's responses are listed in tabular format on the following page.

Table B-4. 1996 SFRP & GSRP Associate Responses to SRP Evaluation

	Arnold	Brooks	Edwards	Eglin	Griffis	Hanscom	Kelly	Kirtland	Lackland	Robins	Tyndall	WPAFB	average
# res	6	48	6	14	31	19	3	32	1	2	10	85	257
1	4.8	4.4	4.6	4.7	4.4	4.9	4.6	4.6	5.0	5.0	4.0	4.7	4.6
2	5.0	4.6	4.1	4.9	4.7	4.7	5.0	4.7	5.0	5.0	4.6	4.8	4.7
3	4.5	4.4	4.0	4.6	4.3	4.2	4.3	4.4	5.0	5.0	4.5	4.3	4.4
4	4.3	4.5	3.8	4.6	4.4	4.4	4.3	4.6	5.0	4.0	4.4	4.5	4.5
5	4.5	4.3	3.3	4.8	4.4	4.5	4.3	4.2	5.0	5.0	3.9	4.4	4.4
6	4.3	4.3	3.7	4.7	4.4	4.5	4.0	3.8	5.0	5.0	3.8	4.2	4.2
7	4.5	4.4	4.2	4.8	4.5	4.3	4.3	4.1	5.0	5.0	4.3	4.3	4.4
8	4.5	4.6	3.0	4.9	4.4	4.3	4.3	4.5	5.0	5.0	4.7	4.5	4.5
9	4.7	4.5	4.7	4.5	4.3	4.5	4.7	4.3	5.0	5.0	4.1	4.5	4.5
10	4.2	4.4	4.7	4.4	4.1	4.1	4.0	4.2	5.0	4.5	3.6	4.4	4.3
11	3.8	4.1	4.5	4.0	3.9	4.1	4.0	4.0	3.0	4.0	3.7	4.0	4.0
12	5.7	4.7	4.3	4.9	4.5	4.9	4.7	4.6	5.0	4.5	4.6	4.5	4.6
Numbers below are percentages													
13a	83	90	83	93	87	75	100	81	100	100	100	86	87
13b	100	89	83	100	94	98	100	94	100	100	100	94	93
14	83	96	100	90	87	80	100	92	100	100	70	84	88
15	17	6	0	33	20	76	33	25	0	100	20	8	39
16a	-	26	17	9	38	23	33	4	-	-	-	30	
16b	100	33	-	40	-	8	-	-	-	-	36	2	
16c	-	41	83	40	62	69	67	96	100	100	64	68	
16d	-	-	-	-	-	-	-	-	-	-	-	0	
17a	-	33	100	17	50	14	67	39	-	50	40	31	35
17b	-	21	-	17	10	14	-	24	-	50	20	16	16
17c	-	-	-	-	10	7	-	-	-	-	-	2	3
17d	100	46	-	66	30	69	33	37	100	-	40	51	46

4. 1996 USAF LABORATORY HSAP MENTOR EVALUATION RESPONSES

Not enough evaluations received (5 total) from Mentors to do useful summary.

5. 1996 HSAP EVALUATION RESPONSES

The summarized results listed below are from the 113 HSAP evaluations received.

HSAP apprentices were asked to rate the following questions on a scale from
1 (below average) to 5 (above average)

1. Your influence on selection of topic/type of work.
2. Working relationship with mentor, other lab scientists.
3. Enhancement of your academic qualifications.
4. Technically challenging work.
5. Lab readiness for you: mentor, task, work plan, equipment.
6. Influence on your career.
7. Increased interest in math/science.
8. Lab research & administrative support.
9. Adequacy of RDL's Apprentice Handbook and administrative materials.
10. Responsiveness of RDL communications.
11. Overall payment procedures.
12. Overall assessment of SRP value to you.
13. Would you apply again next year? Yes (92 %)
14. Will you pursue future studies related to this research? Yes (68 %)
15. Was Tour length satisfactory? Yes (82 %)

	Arnold	Brooks	Edwards	Eglin	Griffiss	Hanscom	Kirtland	Tyndall	WPAFB	Totals
# resp	5	19	7	15	13	2	7	5	40	113
1	2.8	3.3	3.4	3.5	3.4	4.0	3.2	3.6	3.6	3.4
2	4.4	4.6	4.5	4.8	4.6	4.0	4.4	4.0	4.6	4.6
3	4.0	4.2	4.1	4.3	4.5	5.0	4.3	4.6	4.4	4.4
4	3.6	3.9	4.0	4.5	4.2	5.0	4.6	3.8	4.3	4.2
5	4.4	4.1	3.7	4.5	4.1	3.0	3.9	3.6	3.9	4.0
6	3.2	3.6	3.6	4.1	3.8	5.0	3.3	3.8	3.6	3.7
7	2.8	4.1	4.0	3.9	3.9	5.0	3.6	4.0	4.0	3.9
8	3.8	4.1	4.0	4.3	4.0	4.0	4.3	3.8	4.3	4.2
9	4.4	3.6	4.1	4.1	3.5	4.0	3.9	4.0	3.7	3.8
10	4.0	3.8	4.1	3.7	4.1	4.0	3.9	2.4	3.8	3.8
11	4.2	4.2	3.7	3.9	3.8	3.0	3.7	2.6	3.7	3.8
12	4.0	4.5	4.9	4.6	4.6	5.0	4.6	4.2	4.3	4.5
Numbers below are percentages										
13	60%	95%	100%	100%	85%	100%	100%	100%	90%	92%
14	20%	80%	71%	80%	54%	100%	71%	80%	65%	68%
15	100%	70%	71%	100%	100%	50%	86%	60%	80%	82%

BIOGEOCHEMICAL ASSESSMENT OF NATURAL ATTENUATION OF JP-4 CONTAMINATED GROUND WATER IN THE PRESENCE OF FLUORINATED SURFACTANTS

Audrey D. Levine
Associate Professor
Department of Civil and Environmental Engineering
Utah State University
Logan, UT 84322-8200

Final Report for:
Summer Faculty Research Program
Armstrong Laboratory
Tyndall Air Force Base

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, DC

and

Armstrong Laboratory

September 1996

BIOGEOCHEMICAL ASSESSMENT OF NATURAL ATTENUATION OF JP-4 CONTAMINATED GROUND WATER IN THE PRESENCE OF FLUORINATED SURFACTANTS

Audrey D. Levine
Associate Professor
Department of Civil and Environmental Engineering
Utah State University
Logan, UT 84322-8200

Abstract

The biogeochemistry of natural attenuation of petroleum-contaminated ground water was investigated in a field study. The focus of the study was a fire training site located on Tyndall Air Force Base in Florida. The site had been used by the Air Force for about 11 years in fire fighting exercises. An on-site above-ground tank of JP-4 provided fuel for setting controlled fires for the exercises. Various amounts of water and aqueous film forming foams (AFFF) were applied to extinguish the fires. The sources of contamination included leaks from pipelines transporting the fuel, leaks from an oil/water separator, and runoff and percolation from the fire fighting activities. Previous investigations had identified jet fuel contamination at the site, however no active remediation efforts have been conducted to date. The goal of this study was to use biogeochemical monitoring data to delineate redox zones within the site and to identify evidence of natural attenuation of JP-4 contamination. Due to the time constraints of the study, monitoring wells already existing on the site were used for ground water sampling. Four sets of grab samples were collected and analyzed for inorganic and organic water quality parameters. Specific chemical derivatization tests were conducted to provide qualitative evidence of the presence of biological metabolites within various redox zones. In addition to identifying several hydrocarbon metabolites, fluorinated surfactants (AFFF) were detected down-gradient of the hydrocarbon plume. The results of this study provide a frame-work for follow-up modeling and field studies to evaluate the fate, transport, and natural attenuation of JP-4 components and metabolites in the presence of AFFF.

BIOGEOCHEMICAL ASSESSMENT OF NATURAL ATTENUATION OF JP-4 CONTAMINATED GROUND WATER IN THE PRESENCE OF FLUORINATED SURFACTANTS.

Audrey D. Levine

Introduction

Over the past decade significant research has been conducted to evaluate the fate, transport, and environmental and health risks associated with ground water contamination. Recently there has been increased interest in promoting the use of passive remediation processes based on natural biogeochemical attenuation of contaminants. The major objectives of this study were to identify biogeochemical indicators of natural attenuation of petroleum hydrocarbon contaminants under field conditions. The focus of the study was to evaluate relationships between redox conditions and the presence of metabolic byproducts of alkyl benzene degradation at a field site under quasi-steady-state conditions.

Background

Natural attenuation of petroleum hydrocarbons in stationary phase and dissolved plumes has been demonstrated in a number of field and laboratory studies (1-38). Attenuation mechanisms encompass physical dilution, physicochemical sorption and ion exchange, chemical dissolution/precipitation or complexation, and microbial metabolic processes. Key issues influencing the rate and extent of natural attenuation of contaminated ground water include contaminant hydrogeochemistry in conjunction with the availability of subsurface electron acceptor processes, pH, temperature, site geochemistry and hydrology. Modeling and prediction of the rate of natural attenuation and the fate of metabolic by-products is hampered by the lack of field data that integrates geochemical data with contaminant degradation and by-product formation. A brief review of the major factors relevant to determination of dominant redox reactions and a summary of field evidence of metabolic by-product formation is given below.

Redox zones

Redox zones have been characterized in a variety of contaminated aquifers including down-gradient of fuel spills (2,3,7,14,15,16,18,27,35,36,37,38), and landfill leachate plumes (8,10,25, 34). It is widely reported that pH and redox buffering are major controls on biogeochemical reactions in contaminant plumes. A summary of the major types of redox reactions that occur in ground water is given in Table 1 with the standard Gibbs Free Energy ΔG° values. The compound CH_2O is used to refer to a generic organic compound. The actual free energy values depend on the chemical composition of the organic substrate(s), and the concentrations of reactants and products present at a specific location.

Table 1. Summary of major oxidation-reduction reactions that occur in ground water[a]

Type of reaction	Reaction	ΔG^{o} (W), kcal/mol
Methanogenic	$2CH_2O$ ---->CH_3COOH---->CH_4 + CO_2	-22
Sulfate reduction	$2CH_2O$ + SO_4^{-2} + H^+----> $2CO_2$ +HS^- +$2H_2O$	-25
Ferric iron reduction	CH_2O + $4Fe(OH)_3$ + $8H^+$ ---> CO_2 + $4Fe^{+2}$ +$11H_2O$	-28
Manganic reduction	CH_2O + $2MnO_2$ + $4H^+$ ---> CO_2 + $2Mn^{+2}$ + $3H_2O$	-81
Denitrification	$5CH_2O$ + $4NO_3^-$ + $4H^+$ ---> $5CO_2$ + $2N_2$ + $7H_2O$	-114
Oxygen reduction	CH_2O + O_2 ---> CO_2 + H_2O	-120

[a] Adapted from References 7,8,10,13,25,28,34,35

In general, if oxygen is present, biogeochemical reactions tend to be dominated by aerobic reactions. Due to the limited availability of natural mechanisms to replenish oxygen supplies in the subsurface, ground water that contains degradable organic material is likely to be depleted in dissolved oxygen (7,8,10,13,25,28,34,35). Under anoxic conditions electron acceptors such as nitrate, sulfate, Mn(IV), Fe(III) and CO_2 are reduced as organic compounds are oxidized resulting in changes in the redox conditions in the aquifer and increases in concentration of reduced aqueous species such as sulfides,

Fe(II), methane, and ammonia. In contaminated aquifers, the sequence of the reactions down-gradient of a contaminant plume typically follows the sequence given in Table 1 and has been observed in many aquifers with variations depending on the local geochemistry (2,3,7,14,15,16,18,27,35,36,37,38). Some overlap occurs along the interface of successive redox zones.

In sandy aquifers, manganese reduction is a minor electron accepting process and Fe(III) is a major contributor to the oxidation capacity(2,3,7,14,15,16,18,27,35,36,37,38). When Fe(III) is available, iron reduction has been reported to out-compete sulfate reduction and methanogenesis in mediating the oxidation of organic matter (11,36). High concentrations of dissolved iron develop if there is minimal sulfate reduction because generation of sulfide results in the formation of iron sulfide precipitates (11). Methanogenic and sulfate reducing populations compete for the same substrates (acetate and hydrogen) and concurrent methanogenesis and sulfate reduction have been reported in some cases (35). Methogenesis is inhibited by acidic pH levels and low temperatures (4). Fermentation products such as organic acids accumulate in the absence of terminal electron accepting processes.

Due to the dynamic nature of redox processes, it is difficult to measure redox potential directly in the subsurface (34). At a fixed point in the subsurface, changes in the available electron acceptors caused by shifting groundwater flows, surface precipitation events, and seasonal temperature fluctuations result in temporal variations in the dominant terminal electron acceptor processes. In addition, electron acceptors and reduced products of bioreaction can be transported by groundwater convection and may persist down-gradient in zones where there is minimal production of these compounds . Therefore, prediction of local redox reactions from substrate and product measurements and reaction stoichiometries cannot be based on a single measurement.

A detailed look at the reactions listed in Table 1 provides insight into methods for characterizing the operative redox zones within an aquifer. The constituents that are consumed or generated during redox processes include CO_2, CH_4, H_2, NO_3, N_2, SO_4, H_2S, and organic substrate(s). The accumulation of biochemically active elements reflects an imbalance of one or more redox reactions. While no single measurement can provide a true assessment of the operative redox reactions, integrated analyses of the major constituents provides a means of estimating redox conditions. Dissolved hydrogen monitoring has been proposed as a reliable and responsive measure of dominant redox processes (28).

Using the thermodynamic equations given in Table 1, the oxidation reduction potential can be calculated using the Nernst equation. Calculated values of redox potential must be interpreted in the context of site hydrogeochemistry to be of practical value. The relative abundance of oxygen, nitrate, ferrous iron, sulfate, methane, sulfide, and hydrogen in conjunction with ground water hydrodynamics can be used to substantiate and verify redox calculations.

Carbon isotope ratios

An alternative approach to delineate redox zones is to use measurements of dissolved carbon dioxide and carbon isotope ratios (24). Carbon in the environment exists in one of two stable isotopes: ^{12}C and ^{13}C. The ratio of ^{13}C to ^{12}C in dissolved CO_2 is a function of the source of the gas. Isotope ratios are used to evaluate the degree of depletion or enrichment of ^{13}C in a given environment. The standard method for reporting isotope ratios in parts per thousand (‰) is:

$$\delta\,^{13}C\text{–}\sum CO_2 = \left[\frac{\left(^{13}C/^{12}C\right)_{sample}}{\left(^{13}C/^{12}C\right)_{standard}} - 1\right] \times 1000$$

In natural waters the $\delta^{13}C\text{-}\Sigma CO_2$ is controlled by the source and partial pressure of CO_2 and the speciation of dissolved carbon dioxide (2,7,11,24,27,31). In waters with low levels of organic carbon and pH levels below about 6, $\delta^{13}C\text{-}\Sigma CO_2$ is dominated by equilibria between dissolved CO_2 and H_2CO_3 and tends to be more depleted in ^{13}C (34). Photosynthetic reactions selectively utilize $^{12}CO_2$ over $^{13}CO_2$ therefore the residual CO_2 in waters supporting photosynthesis typically reflects isotope ratios that are enriched in ^{13}C. Oils and synthetic chemicals tend to be depleted in ^{13}C, the production of carbon dioxide from contaminant mineralization tends to result in lower $\delta^{13}C\text{-}\Sigma CO_2$ ratios.

Isotope ratios are composite measures of all dissolved carbon in water and therefore, the concentration and composition of dissolved organic compounds in water can be significant. Natural organic matter (NOM) in ground water is derived from biogeochemical reactions and consists of residual heterogeneous, hydrophilic, macromolecular compounds that are operationally defined as humic and fulvic compounds (34). Depending on aquifer geohydrology, NOM can be a significant component of the dissolved organic matter (measured as TOC) in ground water (31,34). Isotope ratios for NOM vary with

its genesis thus posing difficulties in using isotope ratios to differentiate CO_2 produced from contaminant mineralization . Some researchers have evaluated isotope ratios for methane (2). Under anaerobic conditions, ^{12}C is metabolized preferentially to methane and $\delta^{13}C$-ΣCO_2 is enriched while $\delta^{13}C$-ΣCH_4 is depleted in ^{13}C. A summary of reported values of $\delta^{13}C$-ΣCO_2 isotope ratios for various redox conditions is presented in Table 2. The highest isotope ratios are associated with reducing conditions and the lowest ratios are associated with aerobic and nitrate reducing environments. Reported values for methane isotope ratios are around -55 $^{o}/_{oo}$ (2).

Table 2. Comparison of reported values of $\delta^{13}C$-ΣCO_2 in ground water under various redox conditions

Source of CO_2	$\delta^{13}C$-ΣCO_2	Reference
Methanogenic	-11 to +11	2,27
Iron reducing	-6 to +8	7, 27
Near oil lens	-4 to -5	2
Atmospheric CO_2	-8 to -9	27
Anoxic zone	-8 to -9	2
Uncontaminated ground water	-11 to -15	2, 27
Sulfate reducing	-12 to -18	27
Salt marsh	-14 to -17	24
Natural organic matter	-15 to -20	31
Soil CO_2	.-20 to -26	11, 27
Aerobic processes	-19 to -29	7, 27
Nitrate reducing	-19 to -29	30
Oil	-27 to -32	2, 24

Metabolic products of petroleum hydrocarbon degradation

Microbial attenuation of petroleum hydrocarbon plumes has been widely studied. In general, the most soluble and mobile fuel components found in contaminated ground water are benzene, toluene, ethyl benzene, and xylene (BTEX). Degradation of these constituents has been demonstrated in multiple redox environments and relative rates of degradation have been characterized in laboratory and field studies (3,7,14,16,18,35,35,37,38). It has also been reported that n-propyl benzene and 1-methyethyl benzene are conservative within anaerobic plumes and the most stable soluble fuel component in petroleum-contaminated ground water is 1,2,3,4 tetramethylbenzene (15). The recalcitrance of methyl naphthalene has also been observed (26).

Intermediate products of microbial degradation of petroleum hydrocarbons provide biochemical evidence of microbial transformations. Under anaerobic conditions, metabolic by-products include phenols, benzoic acid, one to three methyl benzoic acids and other aromatic acids that are structurally related to alkylbenzene precursors, alicyclic acids, and low molecular weight aliphatic acids (3,5,6,14,17,21,16,33,38). There is a need to determine if metabolites are biologically stable under field conditions and to determine the efficacy of their use as molecular markers of biological contaminant degradation. It is also important to characterize the fate and transport properties of stable metabolites and identify potential health or environmental risks.

A summary of aromatic anaerobic metabolites identified from petroleum hydrocarbon degradation at field sites and in laboratory studies is given in Table 3. Two metabolites of particular interest are: benzyl fumaric and benzyl succinic acids. It has been postulated that benzyl fumaric and benzyl succinic acids are "dead-end" metabolites that might be of significance as biogeochemical indicators of natural attenuation (6). These acids have been identified in sulfate and nitrate reducing environments, although they have not been widely reported as anaerobic metabolites at field sites. The yield of these metabolites is reported as 7 to 10 percent of the mineralized alkyl benzenes (5,6,21). It has been postulated that benzyl succinic acid is microbially dehydrogenated to benzyl fumaric acid (5); therefore the ratio of the two acids is likely to be related to site-specific factors such as electron acceptor, pH, temperature, and other microbial growth requirements.

Table 3. Aromatic anaerobic metabolites identified from petroleum hydrocarbon degradation

Site	Major metabolites	Reference
Methanogenic		
Bemidji, MN	toluic acid dimethyl benzoic acid trimethyl benzoic acid phenyl acetic acid methyl phenyl acetic acid	14
Bordon aquifer, Ontario	2-methyl benzoic acid	3
Laboratory study of toluene and xylene degradation	benzaldehyde benzoate p-Cresol	17
Iron reducing		
Traverse City, MI	dimethyl benzoic acid methyl benzoic acid cresol	38
Sulfate reducing		
Bemidji, MN	toluic acid trimethyl benzoic acid dimethyl benzoic acid	14
Traverse City, MI	methyl benzoic acid cresol	38
Seal Beach, CA; in situ degradation of BTEX (2 to 3 μM BTEX; 0.16mM Sulfate) 60 day duration	benzyl fumaric acid benzyl succinic acid	6
Laboratory study of toluene and xylene degradation	benzyl fumaric acid benzyl succinic acid p-toluic acid	5
Nitrate reducing		
Bemidji, MN	toluic acid trimethyl benzoic acid dimethyl benzoic acid	14
Laboratory column study of alkylated benzene degradation	methyl-benzyl alcohol o-cresol	26
Laboratory study of toluene degradation using pure cultures (0.5 to 1 mM toluene; 5 mM nitrate)	benzyl fumaric acid benzyl succinic acid benzaldehyde benzoate	33
Laboratory: 0.5 to 1 mM m-xylene degradation (5 mM nitrate)	3-methyl benzaldehyde 3-methyl benzoate	33
Laboratory: 1 mM toluene and xylene degradation (5 mM nitrate)	benzyl fumaric acid benzyl succinic acid	21
Laboratory study of toluene and xylene degradation	benzyl fumaric acid benzyl succinic acid p-toluic acid	5

Presently, limited field data exists on the concentration, stability, fate, and transport of metabolites in petroleum hydrocarbon plumes. Extrapolation of laboratory data to field conditions is complex due to site specific biogeochemical conditions. In the laboratory studies summarized in Table 3, relatively high concentrations of xylene or toluene (0.5 to 1 mM) (21,33) were used. The field study (6) was a controlled site in which the fate of injected quantities of BTEX of 2 to 3 μM was tracked over a 60 day period. Contaminant concentrations and other biogeochemical factors vary over several orders of magnitude at field sites and therefore the relative abundance and long-term stability of these compounds remains to be established. If these compounds are stable in the environment it is reasonable to assume

that they should be present in bioactive hydrocarbon plumes under sulfate or nitrate reducing conditions. One reason for the limited field data on benzyl fumaric and benzyl succinic acids relates to analytical limitations. Detection and identification is only possible by using a derivatization procedure to form methyl esters of the acids that can be quantified by GC/MS. Since the derivatization procedure is not part of routine ground water characterization tests, the presence of these metabolites in petroleum hydrocarbon plumes has not been assessed.

Research methodology

The purpose of this project was to conduct a biogeochemical assessment of a petroleum hydrocarbon contaminated ground water to delineate quasi-steady-state redox zones and try to identify evidence of natural attenuation and biological degradation of contaminants. The study is based on using available data from previous and parallel investigations to assess electron acceptor depletion, reduced product formation, intermediate production, and carbon isotope ratios. For this study, pH, conductivity and profiles of SO_4, CO_2, CH_4, Fe (II), organic acids, hydrocarbons, and metabolites were monitored to investigate natural attenuation. Other parameters were estimated using geochemical analyses and existing data.

Site characteristics

This site used for this case study is located at the east side of the flight line at Tyndall Air Force Base (TAFB). TAFB is located at 30 degrees north latitude in the central part of the Florida Panhandle on the western flank of the Appalachicola Embayment in Bay County, Florida. The base is on a peninsula that extends along the shoreline of the Gulf of Mexico. Warm, humid, semitropical conditions are prevalent for about half of the year with convective storms and hurricanes influencing weather. Average annual temperature is about 69° F with lows of 46° F and highs of around 90° F. The mean annual precipitation is 55 inches with about 125 days of recordable precipitation (mostly between June and September). Precipitation percolates directly into the ground or flows into adjacent water bodies. The depth to the surficial ground water ranges from 2 to 10 ft and yearly fluctuations in ground water level of about 5 ft are typical. The aquifer consists of clean, fine-grained quartz and clayey sandy soils that are nearly level, poorly to moderately drained and extend to depths of 80 inches or more (9,20,22,23, 32).

The upper sediments underlying TAFB are sands and gravels about 100 ft thick that comprise the upper portion of the surficial aquifer. The highest ground is about 30 ft above mean sea level. The lower Floridian Aquifer consists of limestones and dolomites. The top of the Floridian Aquifer is 250 below sea level. The aquifer is 1100 ft thick and potable water is derived from the upper 250 to 500 ft of the aquifer (500 to 750 ft below sea level) (9,20,22,23, 32).

The site is a decommissioned fire training area that had been used by the Air Force for about 11 years in fire fighting exercises (1981 to 1992). The site is a flat open grassy area. Fires were set in a pit consisting of a cleared, bermed 0.33 acre area containing an old aircraft or simulated aircraft. The fires were set using "contaminated" JP-4 stored in a 12,000 gallon steel above ground storage tank that was mounted on a concrete pad surrounded by a 3 ft high containment system. The fuel (JP-4) was pumped from the tank through an adjacent pump house and directed to the fire training pit through an underground distribution system. The pit was about 13 ft west of the pump house and was surrounded by a nonvegetated fire prevention zone consisting of shell and sand. Fires were extinguished using water in conjunction with various formulations of aqueous film forming foams (AFFF) consisting of fluorinated surfactants. The sources of contamination include leaks from pipelines transporting the fuel, leaks from an oil/water separator, overspill of fuel in the pit during exercises, and runoff and percolation from the fire fighting activities.

Three ground water and soil investigations have been conducted as part of the Air Force Installation Restoration Program and consequently thirteen monitoring wells are currently distributed across the site. Twelve of the wells are shallow wells with screened intervals between 2 and 15 ft below the ground surface. In 1986, three monitoring wells were installed : T11-1 (upgradient), and T11-2 and T11-3 (down-gradient). In 1988, two additional upgradient wells were installed (TY22FTA and TY23FTA). In 1991, two shallow wells were installed within the contaminant plume: AFMW-1(near the above ground storage tank) and AFMW-2 (near the oil/water separator). A final set of shallow monitoring wells was installed in 1993 to define the horizontal (MW-1 through MW-5) and vertical (DMW-1 to 37 ft) extent of hydrocarbon contamination at the site. The wells were installed using the hollow-stem auger method of drilling and consist of 10 foot sections of 2 inch diameter PVC screen

attached to PVC casing. Wells are sealed with bentonite and cement grout and protected by steel casings that are embedded in concrete pads at each well head (9,20,22,23,32).

A plan view of the site is shown in Figure 1. The land slopes gently towards Little Cedar Bayou. Structures at the site included a fuel storage tank, a lined training pit where fires were set and extinguished, an oil/water separator that discharges wastewater to Little Cedar Bayou. A storm water drain and outfall are located to the east of the oil/water separator and drainfield. Soil vapor analyses in conjunction with ground water sampling were used to characterize the site. No estimate of the mass of contamination at the site is available at present. A small plume was delineated at the south side of the pump house that encompasses the above ground storage tank and pump house (MW-5). This plume contained free phase product up to 3 ft in thickness in 1994. A second larger plume that plume contained about 0.5 inches of free product was observed along the distribution piping and under the fire training pit (AFMW-1). The vertical extent of the contamination was assessed by a single deep well (DMW-1). No petroleum hydrocarbons were detected in the deeper well (9,20,22,23,32).

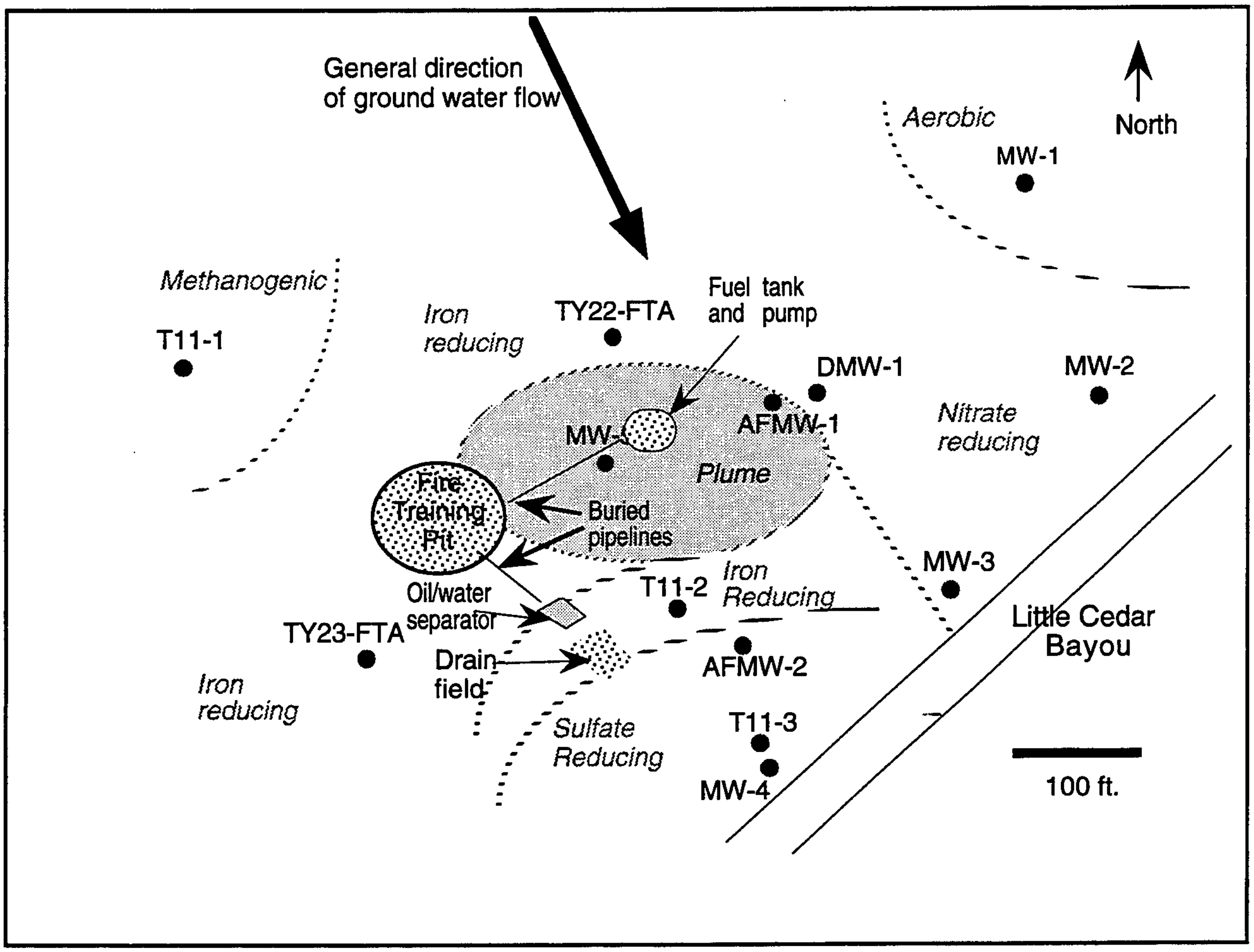


Figure 1. Schematic of the Fire Training Site at Tyndall Air Force Base. Approximate location of fire training pit, fuel storage tank, underground pipelines, monitoring wells, oil/water separator, and drainfield are shown. The general location of redox zones is also indicated (adapted from reference 32).

The hydraulic gradient at the site was evaluated in previous studies using water table elevation measurements in existing wells. Slug tests in wells were used to estimate hydraulic conductivity. In general, hydraulic gradient and conductivity vary over the site. Upgradient of the plume the hydraulic conductivity ranges from 0.881 (T11-1) to 1.781 (TY22FTA) ft/day. Down-gradient of the plume the hydraulic conductivity is about 0.348 (T11-3) ft/day. Near the Little Cedar Bayou hydraulic conductivity has been reported as 1.274 ft/day (MW-1) and 0.091 ft/day (MW-2). The slope of the water table is

relatively shallow with a hydraulic gradient of about 0.01 to 0.03 ft/ft.. The ground water velocity has been estimated to range from 0.03 to 0.112 ft/day (9,20,22,23,32).

Sampling

Due to the time constraints of the project, the sampling program was based on sampling from 13 existing wells on the site. For the purposes of this study it was assumed that each well represented quasi-steady-state well-mixed conditions reflective of the region of the well. Existing data from the site were reviewed to determine appropriate sampling strategies. Four sampling events were conducted during the summer of 1996. For each sampling event, each well was pre-bailed using a bottom discharge bailer. Samples were collected with the bailer and either transferred to sampling containers; filtered, and transferred to sampling containers; or used for on-site analyses. On site measurements of pH, conductivity, temperature and dissolved oxygen were made using field calibrated probes.

Analyses

The analytical methods used were based on Standard Methods. pH, conductivity, dissolved oxygen, and temperature were measured in the field using calibrated probes. Samples for analyses for anions, total organic carbon, and ultraviolet absorbance were preserved at 4° C and usually completed within 24 hours of sample collection. For analysis of ferrous iron, samples were poured immediately after collection into pre-washed 5-mL syringes (to prevent oxidation of iron), and a measured volume was filtered (0.45 µm pore size) into glass vials containing 2 mL of Ferrozine-Herpes buffer solution. The iron concentration was determined spectrophometrically by reading absorbance at 652 nm and comparing readings to a standard curve. Anions (nitrate, chloride, bromide and sulfate) were measured using a Dionex ion chromatography system with isocratic sodium hydroxide eluent and quantified by conductivity. Total organic carbon was measured using a Shimadzu Total Organic Carbon analyzer. Ultraviolet absorbance was measured using a Cary UV-VIS spectrophotometer. Surfactant levels were estimated using a Hach chloroform extraction to determine methylene blue active substances (MBAS). Samples for measurement of DIC, methane, and $\delta^{13}C$ were collected and analyzed by Glynnis Bugna from the Department of Oceanography at Florida State University. Isotope samples were transferred to 10 mL syringes and filtered into evacuated vials. Sample vials were pressurized to ambient pressure with nitrogen gas before they were analyzed for DIC and $\delta^{13}C$ with an IR/GC mass spectrometer. All analyses were conducted using standardized QA/QC protocols.

Metabolite derivitization and analysis.

Metabolite sample collection, processing and analysis followed the procedure outlined by Bellar et al. (5,6). Samples were collected in 1 L glass bottles and immediately acidified to pH 1 using HCl. Each sample was spiked with 0.1 µM 4-fluorobenzoic acid to track the efficiency of the extraction. Standards of benzyl succinic acid in water were run in parallel with the field samples to verify the derivitization procedure. Samples were extracted 3 times with high purity diethyl ether using liquid-liquid extraction in 2 L separatory funnels. Extracted samples were rotary evaporated to 2 to 5 mL, dried with precleaned Na_2SO_4. Dried samples were derivatized with ethereal diazomethane and exchanged into high purity dichloromethane using high purity N_2 at room temperature. Samples were spiked with crysene as an internal standard and analyzed with a GC/MS DB-5 fused silica capillary column. Internal standard quantification was used to evaluate GC/MS response factors. Fluorinated surfactants were also determined by this derivitization and analysis procedure.

Analysis of hydrocarbons

Samples for hydrocarbon analysis were collected in 40 mL VOA vials with hole caps and Teflon-faced septa. During sample collection, the vials were allowed to overflow to eliminate headspace, preserved with HCl and capped with septa. Hydrocarbon analyses were conducted using a Solid Phase Micro Extraction (SPME) and analyzed using gas chromatography with flame ionization detection. For the SPME extractions, samples were poured into a calibrated extraction vial to the 35 mL mark, capped with Teflon-faced caps, and stirred. The sample was then spiked with 10-µL of an internal standard solution which consisted of 342-µg/mL d10-ethylbenzene in 2-propanol. An extraction holder containing a pre-calibrated SPME fiber coated with 100-µm of polymethylsiloxane (Supelco, Inc) was adjusted to provide 1-cm exposure. The septum of the extraction vial was pierced with the fiber holder needle and the fiber was extended into the sample headspace for 20 minutes at room temperature. After the exposure period, the SPME fiber was withdrawn, and the fiber was injected immediately into a gas chromatograph (HP-5890 Series II) with a flame ionization detector. A split/splitless injection port (250°C) was used in the splitless mode and was purged at 3-min. The SPME fiber remained extended into the injection port for 20-min to ensure there was no carry-over even if a heavily contaminated water sample was analyzed. The

GC oven was pre-cooled to -10°C before and during the injection procedure using liquid nitrogen. The oven was set initially to -10° C with an initial isothermal hold of 3-min., followed by a linear temperature program of 10° C/min to a final temperature of 250°C and a 6-min final isothermal hold. The chromatographic separations were performed with a fused silica capillary column, (30-m by 0.25-mm), and coated with 1.0-μm of a 5%-phenyl substituted polymethylsiloxane (DB-5) bonded and crosslinked stationary phase. Helium was used as the carrier gas with a constant head pressure of 15-psig. Calibration curves were developed from standard solutions of benzene, toluene, ethylbenzene, 1,4-dimethylbenzene, isopropylbenzene, n-propylbenzene, butylbenzene, and 2-methylnaphthalene at various concentrations in 2-propanol. The standard solutions also contained d_{10}-ethylbenzene in the same concentrations as the internal standard spiking solution used with the groundwater samples. Standard extractions were performed by spiking 35-mL samples of distilled/deionized water with 10-μL aliquots of the standard solutions.

Organic Acid analysis

Organic acid analyses was conducted using HPLC (19). Samples for organic acid analysis were filtered and acidified immediately after sample collection. Samples were analyzed using a BioRad Aminex ion exclusion column HPX-87H (300 mm by 7.8 mm) column. The mobile phase was 0.013N sulfuric acid at a flow rate of 0.6 mL/min . Peaks were detected at 210 nm and identified by comparing retention times of unknowns with retention times for standard organic acids.

Results

The major contaminants of concern at the fire training site are petroleum hydrocarbons and monitoring data has been collected sporadically since 1988 in conjunction with various studies (9,20,22,23,32). Variations in BTEX and methyl-naphthalene concentrations at the two wells within the plume (MW-5 and AFMW-1) are presented in Figures 2 and 3. Data from a down-gradient well (T11-3) are shown in Figure 4. As shown, significant fluctuations have occurred in the reported levels for BTEX in each well over the 8 year monitoring period with a general decrease in contaminant levels since the 1994 sampling. The site has not been used for fire training since 1992 and the storage tank has been removed eliminating additional sources of contamination outside of the existing plumes.

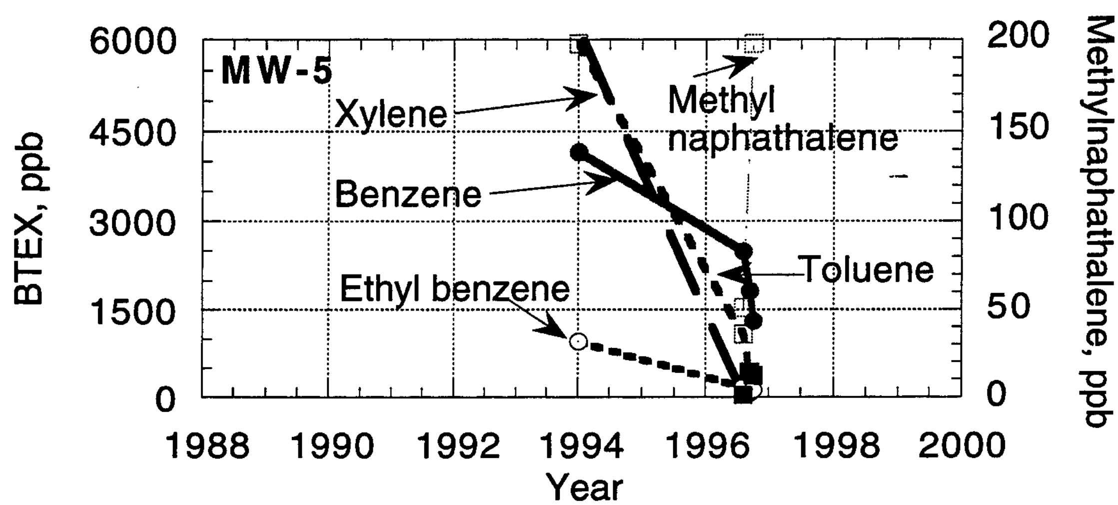


Figure 2. Summary of BTEX and methyl-naphthalene monitoring data for well MW-5.

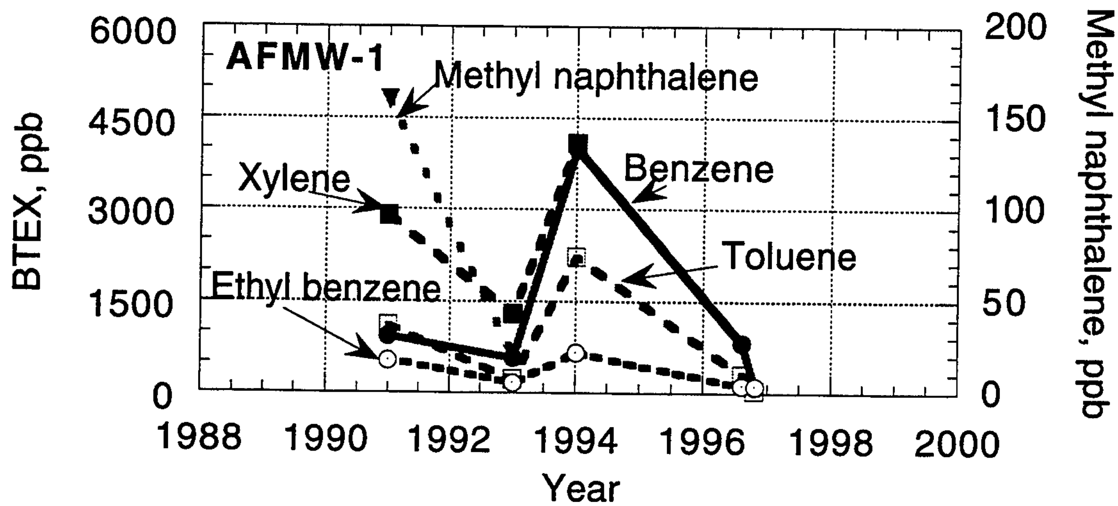


Figure 3. Summary of BTEX and methyl-naphthalene monitoring data for well AFMW-2.

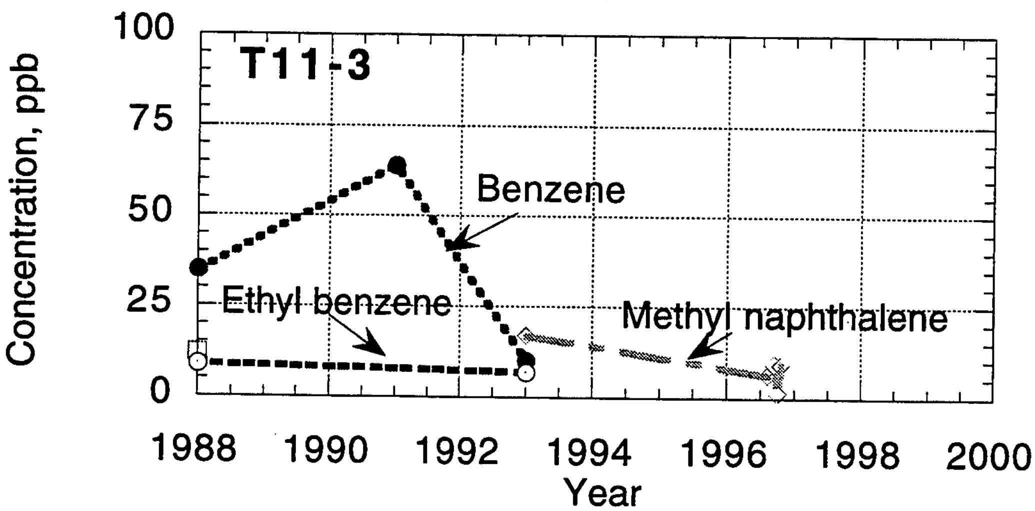


Figure 4. Summary of BTEX and methyl-naphthalene monitoring data for well-T11-3.

Water Quality Analysis

For preliminary assessment of the geochemistry of the site it was divided into four general regions with respect to the hydrocarbon plume: up-gradient; within the hydrocarbon plume, down-gradient; and down-gradient and outside of the zone of influence of the plume. A summary of general water quality characteristics from each zone is given in Table 4. As shown, there is significant variability in the data. Several of the wells down-gradient of the plume had strong sulfide odors, however sulfate was not detected in any of the samples. It is likely that sulfate was reduced at the same rate as it became bioavailable and therefore was below detection limits even in the sulfate reducing zone (11,36). The wells that are upgradient and within the plume generally reflect anaerobic conditions and the down gradient wells tend be more iron or sulfate reducing. The wells that are near the bayou and outside of the zone of influence display nitrate reducing or aerobic conditions. The pH of the site isgenerally below 6 with slightly higher values in the down-gradient wells. The water temperature ranged between 25 and 30 C (reflecting ambient conditions). The TOC levels across the site are high reflecting significant dissolved natural organic matter. The DIC, methane and carbon isotope values are consistent with levels reported in the literature (see Table 2) for the various redox zones.

Table 4. Summary of water quality data from fire training area wells sampled during summer 1996.

Parameter	Up-gradient wells T11-1;TY22 FTA; TY23FTA		Plume wells MW-5; AFMW1		Down-gradient wells T11-2; T11-3; AFMW-2; MW-4		Outside of zone of influence MW-1; MW-2; MW-3	
	Range	(mean ± std)	Range	(mean ± std)	Range	(mean ± std)	Range	(mean ± std)
pH	5.2 to 6.6	(5.8 ± 0.51)	5.3 to 5.6	(5.5 ± 0.12)	5.6 to 6.8	(6.2 ± 0.5)	4.1 to 5.8	(5 ± .8)
Temp, C	26 to 29	(27.7 ± 1.4)	26 to 28	(26.8 ± 0.7)	24 to 29	(26.8 ± 1.7)	27 to 29	(27.7 ±1.0)
Bromide, ppm	0.1 to 1.2	(0.6 ± 0.4)	0.9 to 1.7	(1.4 ± 0.3)	0.4 to 2.1	(1.1± 0.66)	0.2 to 1.1	(0.7 ± 0.3)
Chloride[a], ppm	6 to17	(12.6 ± 5.9)	20 to 28	(24 ± 6)	4 to 32	(19 ± 9)	34 to 55	(44±15)
Iron (II),ppm	0.2 to 16	(3.9 ± 6.3)	3 to 19	(10 ± 8)	0.4 to 15	(5.6 ±5.9)	0.9 to 14	(5.7 ±4.4)
Nitrate[a], ppm	1 to17	(6.3 ± 6.3)	0.4 to 2	(1.3 ± 0.7)	0.4 to 3	(1.2 ± 1.5)	0.2 to 28	(17.7 ± 12.5)
TOC, ppm	25 to 68	(41 ±16)	72 to 119	(86 ± 22)	29 to 122	(58 ± 25)	2 to 61	(30 ± 21)
DIC, mM	5 to13	(9.4 ± 2.6)	7 to13	(9.9 ± 2.4)	4 to 12	(6.9± 2.5)	2 to 4	(2.7 ± 0.9)
Methane, µM	429 to 764	(637 ± 127)	288 to 651	(501 ± 154)	79 to 525	(201 ± 163)	0.2 to 146	(45± 57)
$\delta^{13}C\text{-}\Sigma CO_2$, ‰	-8.5 to 2.0	(-1.2 ± 4.3)	-2.4 to 1.1	(-0.4 ± 1.6)	-12.1 to -1.4	(-7.1± 4.1)	-9.4 to -20.2	(-14 ± 4.8)

[a] Due to co-elution of sulfide with nitrate and chloride peaks, values for samples containing sulfides were estimated.

Dominant redox zones

Based on assessment of key redox indicators, dominant redox zones for this site were delineated. The key redox intermediates measured in this study were nitrate, iron, and methane. Sulfide levels were estimated from conductivity levels and water quality data. The levels of inorganic redox intermediates (nitrate, ferrous iron, and sulfide) as a function of distance down-gradient are shown in Figure 5 and the levels of methane, inorganic carbon, carbon isotopes, hydrocarbon levels, and TOC are shown in Figure 6. As shown the level of nitrate is depleted within the plume and immediately down-gradient. Iron decreases down-gradient of the plume due to precipitation as iron sulfide, changes in the operative redox processes, or other factors. As shown in Figure 6, methane, DIC, and carbon isotope data are consistent with water quality data.. Methane, dissolved inorganic carbon, and carbon isotope ratios tend to decrease down-gradient of the plume in parallel with reductions in hydrocarbon concentrations. TOC levels throughout the site are quite high due to natural sources of organic matter.

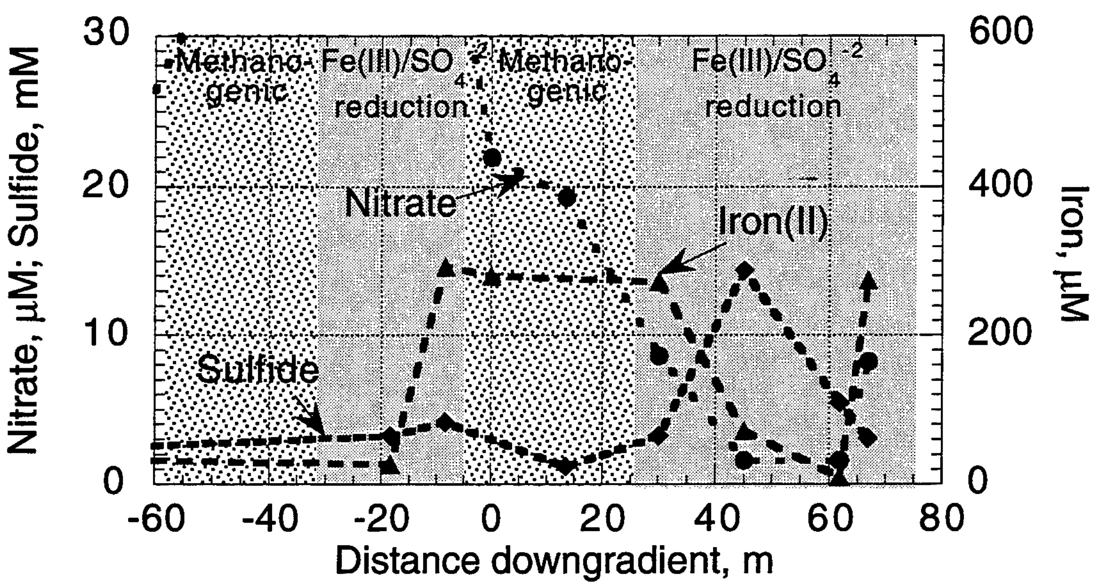


Figure 5. Comparison of redox intermediates (nitrate, ferrous iron, and sulfide) as a function of distance down-gradient at the Fire Training Site (data from summer 1996).

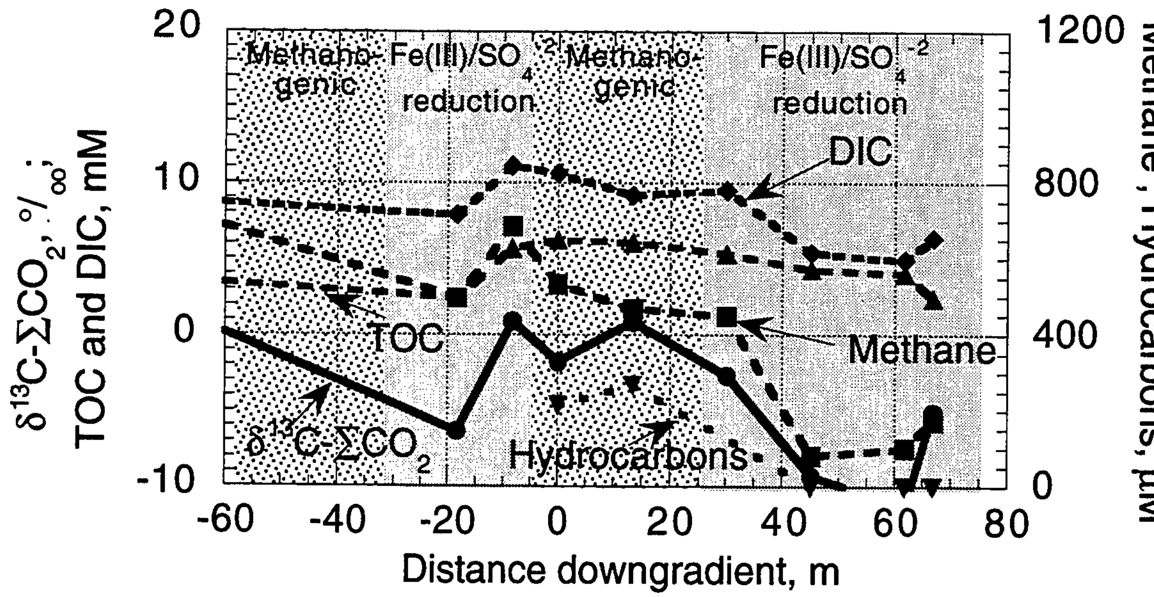


Figure 6. Comparison of dissolved inorganic and organic carbon, carbon isotopes, and hydrocarbon levels as a function of distance down-gradient at the Fire Training Site.

The water quality data were used to determine the dominant redox conditions for each well and relative redox potentials were calculated. A summary of the calculated redox potential as a function of distance down-gradient from the highest zone of contamination (MW-5) is shown in Figure 7 in comparison with the hydraulic gradient and measured pH at each location.

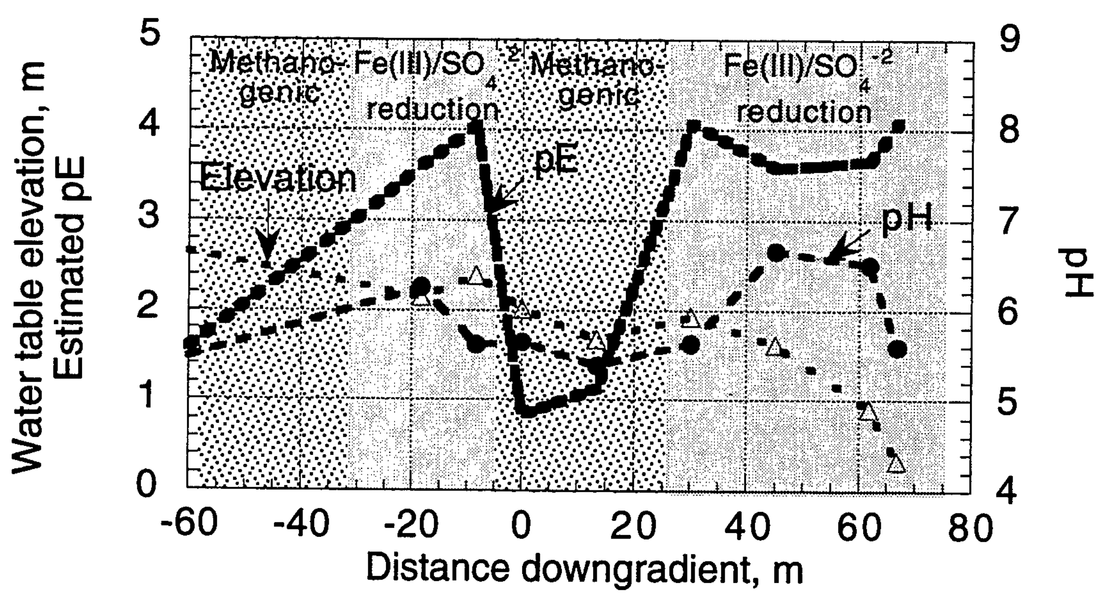


Figure 7. Comparison of calculated pE and measured pH values as a function of distance from the hydrocarbon plume. Hydraulic gradient data are from reference 32.

Contaminant transport

The concentrations of BTEX, methyl naphthalene, isopropyl benzene and n-propyl benzene as a function of distance down-gradient are presented in Figures 8 and 9. As shown, the BTEX levels are nondetectable within about 65 m down-gradient. Consistent with previous findings, isopropyl benzene and n-propyl benzene are not degraded within the methanogenic zone of the plume, but are rapidly degraded in the Fe(III) and sulfate reduction zone.

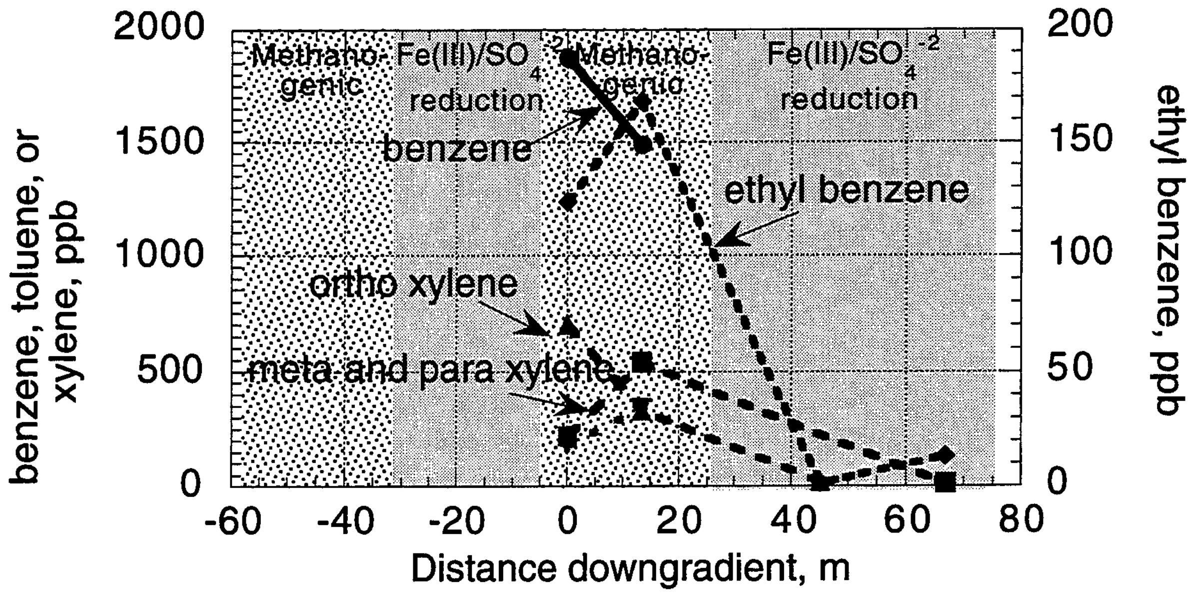


Figure 8. Comparison of benzene, toluene, ethyl benzene and xylenes as a function of distance down-gradient from MW-5. Data are average values from summer 1996 sampling.

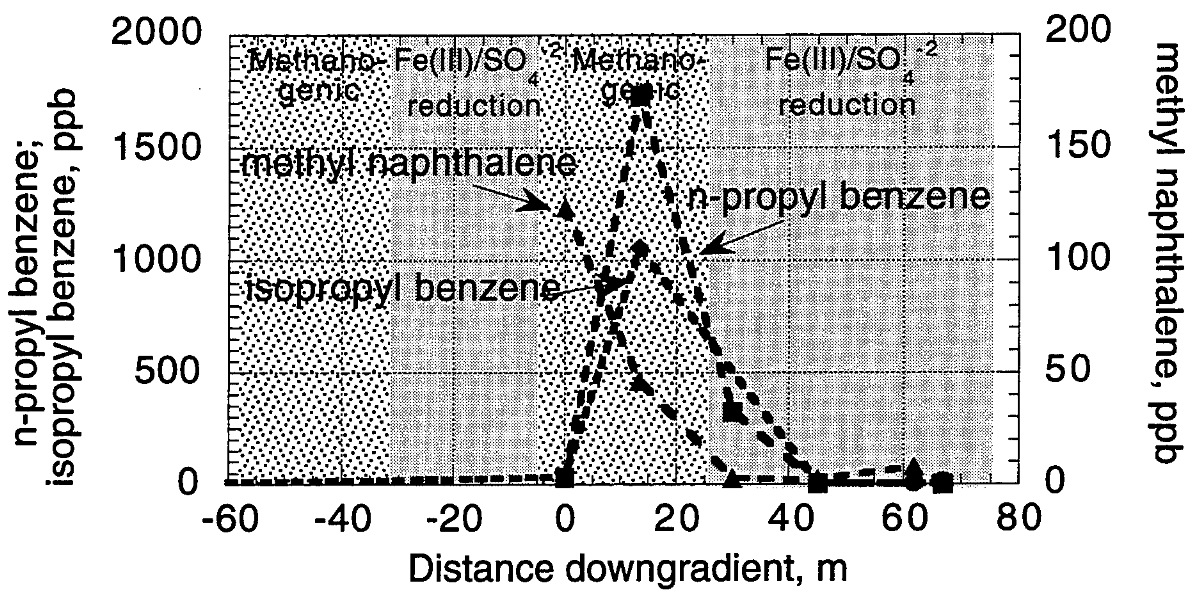


Figure 9. Comparison of n-propyl benzene, isopropyl benzene and methyl naphthalene as a function of distance down-gradient from MW-5. Data are average values from summer 1996 sampling.

Carbon isotope ratios

A comparison of carbon isotope ratios for carbon dioxide and methane as a function of distance down-gradient is shown in Figure 10. The isotope results generally follow the trends reported in Table 2 with higher values associated with methanogenic conditions and lower values for sulfate reduction. However, interference from the background NOM in delineating the zone of contamination using this approach is evident. Down-gradient of the plume, the methane isotope ratio increases while the carbon dioxide ratio decreases reflecting utilization of the hydrocarbon substrate.

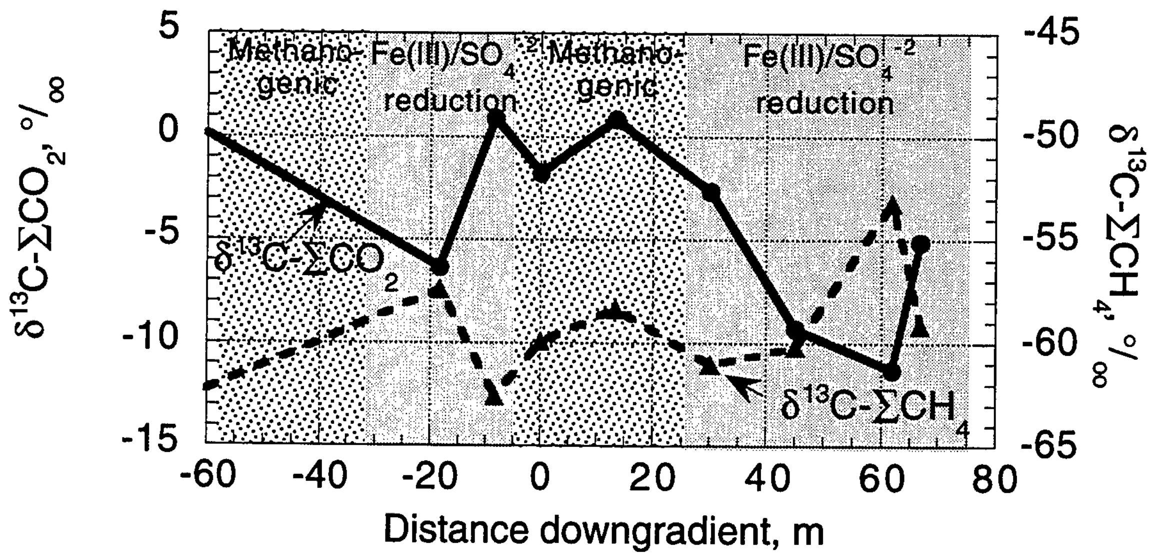


Figure 10. Comparison of carbon dioxide and methane isotope ratios as a function of distance down-gradient from MW-5 (Data from Florida State University, 1996).

Relationships between measured methane concentrations and dissolved organic carbon and carbon isotope ratios are shown in Figure 11. In general there is a linear relationship between methane and the inorganic carbon, however there is a fair amount of scatter in the data most likely due to interferences from NOM.

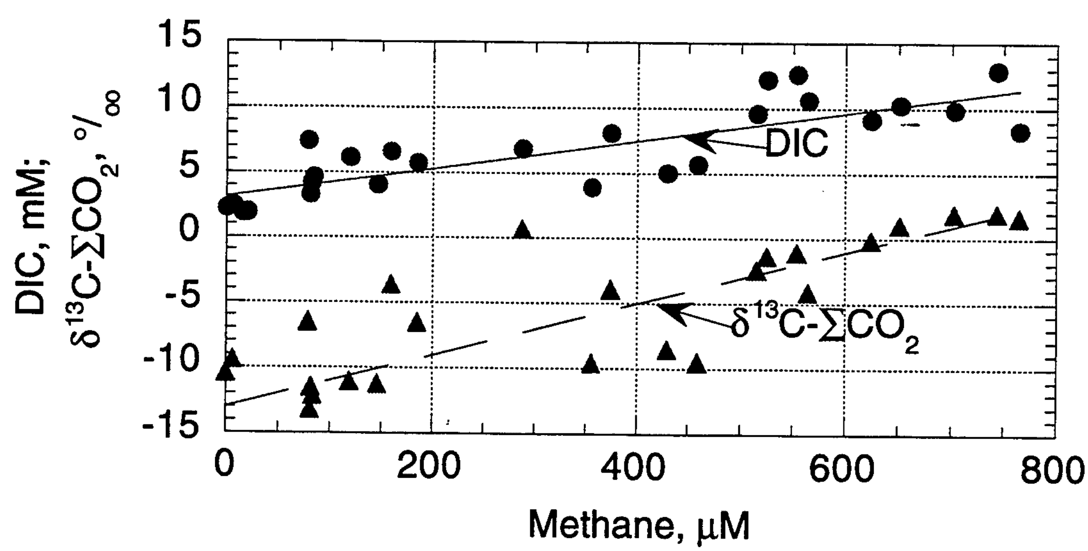


Figure 11. Correlation of methane levels in ground water at the fire training site with levles of dissolved organic carbon (DIC) and carbon isotope ratios. (Data from Florida State University, 1996).

Dissolved organic carbon and metabolite measurements

The shallow ground water at this site contains fairly high levels of NOM that may be of significance in biogeochemical attenuation of the hydrocarbon plume. Some preliminary efforts at characterizing the dissolved TOC were conducted during this study. An analysis of the degree of aromaticity of TOC can provide insight into biogeochemical transformations that occur across the site. The specific ultra-violet absorbance (SUVA) is a measure of the degree of aromaticity of the dissolved organic materials. Since aromatic and double-bonded compounds have strong UV absorbance spectra, the ratio of UV absorbance to TOC (SUVA) provides a means to track aromaticity across the site. A comparison of dissolved organic carbon, SUVA, and surfactant measurements as a function of distance down-gradient is given in Figure 12. As shown, while the TOC values are fairly high across the site, the SUVA decreases indicating a decrease in aromaticity in the iron and sulfate reducing zones down-gradient. Low levels of acetate and formate (below 5 ppm) were detected in all methanogenic, iron and sulfate reducing zones. The upgradient wells contain higher levels of natural organic matter that appears to have higher aromatic content. The down-gradient wells are in close proximity to the Little Cedar Bayou and may be subject to some tidal influence that could modify the concentration and composition of the TOC. The role of the dissolved organic matter in mediating transport, bioavailability, and toxicity of contaminants and redox intermediates in contaminated plumes can not be determined from the present study.

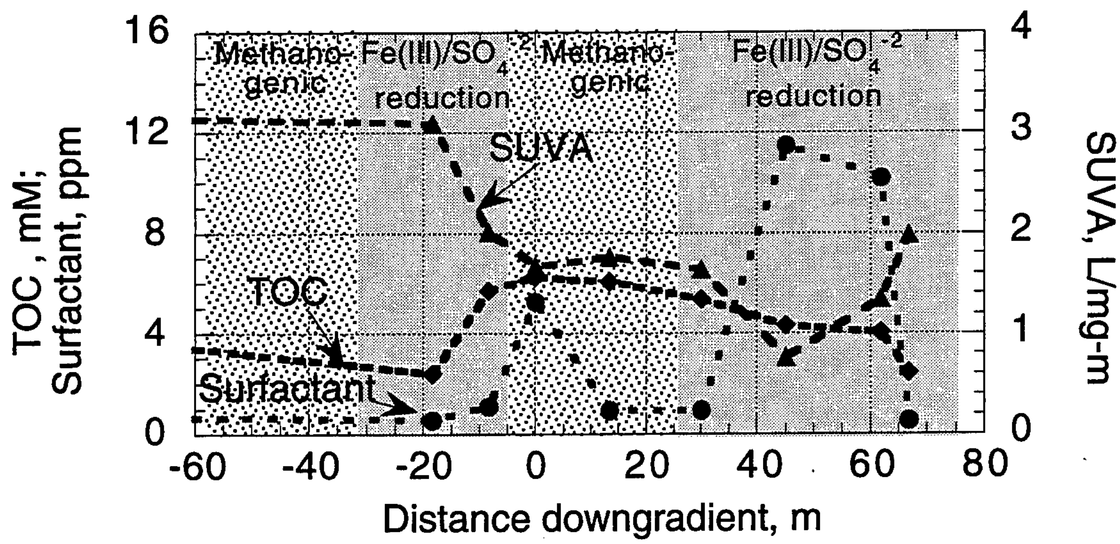


Figure 12. Comparison of dissolved organic carbon (TOC) levels with surfactant estimates and specific UV absorbance (SUVA) as a function of distance down-gradient from MW-5.

It is also interesting to note the surfactant concentration around the site. There should be no natural sources of surfactant in the ground water. However, the AFFF used in fire training appears to be transported in the ground water. The apparent solubility of contaminants and TOC can be increased in the presence of surfactants. In some cases surfactants can reduce the degree of sorption and increase contaminant mobility and reactivity.

Metabolite derivitization and analysis

Multiple samples from each well up-gradient and down-gradient of the contaminant plume were derivatized and analyzed using GC/MS to identify potential metabolites of anaerobic degradation with specific focus on benzyl succinic and benzyl fumaric acids. Based on previous findings (5,6,21,33) the most likely zone for formation of these acids would be down-gradient of MW-5 under sulfate reducing conditions (AFMW-2, T11-3, and MW-4) or nitrate reducing conditions. However neither benzyl succinic nor benzyl fumaric acid were detected in multiple samples from all zones of the site. A summary of the metabolites identified in each zone is given in Table 5. These findings are consistent with data reported from other field sites (14,3,38).

In addition to metabolite evaluation, fluorinated surfactants (AFFF) were detected upgradient of the plume (TY22 FTA), within the plume (MW-5 and AFMW-1), and down-gradient of the plume (T11-2, AFMW-2, T11-3, and MW-4). The transport of AFFF down-gradient of the plume is not suprising due to the large quantities of foam that were applied to the site during its operation in fire training. An unexpected finding, however, was the detection of AFFF in MW-3 which appeared to be outside of the zone of influence of the plume based on all other analyses. The presence of AFFF within the ground water introduces several important questions relating to attenuation processes in progress at this site. Limited information is currently available about transport properties of AFFF in the subsurface. If AFFF are transported more rapidly than the other dissolved constituents, the detection of AFFF outside of the active plume zone (MW-3) may serve as an "early warning" of continued migration of the plume. In addition, the role of AFFF in mediating or inhibiting biogeochemical reactions needs to be elucidated. The absence of benzyl succinic and benzyl fumaric acids in the sulfate reducing zone may be related to the fate and transport of AFFF within the ground water. Alternatively, if AFFF is non-reactive in a biogeochemical context, then the use of benzyl succinic and benzyl fumaric acids as indicators of anaerobic hydrocarbon degradation does not appear appropriate for this site. The higher temperatures, high background levels of TOC, low pH levels, surfactant matix characteristic of this site may have promoted alternative pathways for microbial degradation of hydrocarbons that do not yeild stable forms of benzyl succinic and benzyl fumaric acids.

Table 5. Summary of anaerobic metabolites detected in ground water at fire training site.

Zone	Wells	Metabolites detected
Upgradient methanogenic	T11-1	None
Upgradient Iron and/or sulfate reducing	TY-22FTA; TY23FTA	None
Deep monitoring well (25 ft)	DMW-1	None
Plume	MW-5; AFMW-1	methyl benzoic acid benzene acetic acid
Down-gradient iron reducing	T11-2	benzene acetic acid dimethyl benzoic acid
Down-gradient sulfate reducing	AFMW-2; T11-3; MW-4	trimethyl benzoic acid
Nitrate reducing and/or aerobic (outside plume)	MW-1; MW-2; MW-3	None

Conclusions

This field investigation provided an opportunity to test current theories of biogeochemical attenuation. The findings from this site should be of particular value at other locations where AFFF have been applied in the conduct of Air Force operations.

- The characteristics of this field site containing JP-4 contaminated ground water from a decommissioned Air Force fire training area are consistent with patterns observed at other petroleum-hydrocarbon contaminated sites (1-38). This site tended to display higher temperatures and higher background levels of TOC than are typically reported. These factors play a key role in microbial reaction rates.
- Monitoring of redox intermediates, dissolved inorganic and organic carbon, methane, and carbon isotope ratios provided a basis for estimating dominant redox processes at this site. Carbon isotope data tracked biological changes but did not provide efective "stand-alone" assessmentsof contaminant degradation due to the high background levels of TOC.
- The composition of TOC varies across the site in terms of aromaticity. Further characterizaion of the role of dissolved organic matter in biogeochemical attenuation would be of value.
- Metabolites of anaerobic degradation were identified in methanogenic, iron reducing, and sulfate reducing regions of the site. Benzyl succinic and benzyl fumaric acid were not identified in any of the samples collected from this site. The absence of these metabolites at this site casts doubt upon the efficacy of their use as molecular markers of biological attenuation and should be verified by additional analyses in the zone down-gradient from the hydrocarbon plume.
- Significant levels of AFFF in the ground water were detected at this site. The presence of these compounds may be of significance in biogeochemical attenuation of hydrocarbon contaminants.

Recommendations

To verify the preliminary findings reported in this study, several topics are recommended:

- Seasonal variations in ground water quality and their overall role in reaction rates for natural attenuation should be evaluated by continuing the current sampling program on a quarterly basis and adding the measurement of dissolved hydrogen, ammonia nitrogen, sulfide, and organic acids.
- Verification of the metabolite distribution of the site should also be conducted to develop statistically sound data. Methods to streamline and automate the derivitization procedure should be explored if this test is to be conducted routinely.
- Carbon isotope studies should be continued to evaluate seasonal variations, the influence of rainwater infiltration, and tidal influence on ground water near the bayou. In addition, it is recommended that isotope measurements be made of the free product as well as the dissolved phase in the highly contaminated wells. Further characterization of the influence of ground water NOM on isotope ratios would also be of value.
- Further investigation of AFFF is warranted with respect to fate and transport properties, role in biogeochemical reactions, and risk analysis issues relating to remediation of fire training areas.
- The use of hydrogen monitoring to delineate redox zones in conjunction with other redox intermediates would improve the overall confidence in the redox characterization. It is suggested that methods of sampling dissolved gases in the field and/or using solid phase extraction techniques to stabilize or preserve dissolved hydrogen be developed and tested at this site to facilitate analysis.
- Follow-up studies to characterize and model the ground water geochemistry, biological reaction rates, the fate and transport of AFFF in ground water, and the role of dissolved organic carbon in mediating natural attenuation are recommended.

Acknowledgments

The Armstrong Laboratory staff at Tyndall Air Force Base provided assistance and support for this study. The project supervisor was Dr. Thomas Stauffer. Dr. E. Laurence Libelo provided field and laboratory support. The laborious derivitization, gas chromatography, and interpretation of mass spectra by Dr. Tim Shelley provided results critical to the success of this project. Dr. Howard Mayfield developed SPME procedures for hydrocarbon analyses and conducted gas chromatography and mass spectrometry analyses. Dr. William Rixey (University of Houston) provided assistance with SPME and gas chromatography. Mr. Michael Henley provided assistance and support with liquid chromatography and mass spectrometry. Ion chromatography, TOC analyses and general laboratory assistance by Ms. Eila Burr and Ms. Marlene Cantrell are appreciated. The project would not have been possible without the generous assistance with field sampling and analysis of carbon dioxide, methane, and carbon isotopes by Ms. Glynnis Bugna and her sampling crew from Florida State University; iron analyses by Mr. Eric Henry and field sample collection by Ms. Leslie Buck and Dr. Daniel Smith.

References

1. Acton, D.W., and J.F. Barker (1992), "In situ biodegradation potential of aromatic hydrocarbons in anaerobic groundwaters" *Journal of Contaminant Hydrology*, 9, 325-352.
2. Baedecker, M.J., I.M. Cozzarelli, and R.P. Eganhouse (1993), "Crude oil in a shallow sand and gravel aquifer III. Biogeochemical reactions and mass balance modeling in anoxic groundwater" *Applied Geochemistry*, vol.8, pp 569-586.
3. Barbaro, J.M, J.F. Barker, L.A. Lemon, and CI Mayfield (1992) "Biotransformation of BTEX under anaerobic, denitrifying conditions: Field and laboratory observations" *Journal of Contaminant Hydrology*, 11, 245-272.
4. Beeman, R.E., and J.M. Suflita (1990), "Environmental factors influencing methanogenesis in a shallow anoxic aquifer: a field and laboratory study" *Journal of Industrial Microbiology*, 5, 45-58.
5. Bellar, H.R., M. Reinhard, and D.Grbic-Galic (1992) "Metabolic By-Products of Anaerobic Toluene Degradation by Sulfate Reducing Enrichment Cultures" *Applied and Environmental Microbiology*, Vol.58, No. 9., 3192-3195.

6. Bellar, H.R., W-H Ding, and M. Reinhard, (1995), "Byproducts of Anaerobic Alkylbenzene Metabolism Useful as Indicators of in Situ Bioremediation" *Environmental Science and Technology*, 29, 11, 2864-2870.

7. Bennett, P.C., D.E. Siegel, M.J. Baedecker, and M.F. Hult (1993), "Crude oil in a shallow sand and gravel aquifer I. Hydrogeology and inorganic geochemistry" *Applied Geochemistry*, vol.8, pp 529-549.

8. Bjerg, P.L., K.Rugge, J.K. Pedersen, and T.H. Christensen (1995), "Distribution of Redox Sensitive Groundwater Quality Parameters Down-gradient of a Landfill, *Environmental Science and Technology*, vol.29 no.5, 1387-1394.

9. CH2M Hill (1981), "Installation Restoration Program Records Search for Tyndall Air Force Base, Florida, Phae I, Stage I of the Installation Restoration Program" Tyndall AFB, Florida.

10. Chapelle, F.H. (1993) *Ground Water Microbiology and Geochemistry*. John Wiley and Sons, New York.

11. Chapelle, F.H. and D.R. Lovely (1992), "Competitive Exclusion of Sulfate Reduction by Fe(III) reducing bactera: A mechanism for producing discrete zones of high-iron ground water". *Ground water*, vol.30, no.1, 29-36.

12. Chapelle, F.H. and P.B. McMahon (1995), "Geochemistry of dissolved inorganic carbon in a Coastal Plain aquifer. 1. Sulfate from confining beds as an oxidant in microbial CO_2 production". *Journal of Hydrology*, 127, 75-108..

13. Cozzarelli, I.M., J.S. Herman, and M.J. Baedecker (1995), "Fate of Microbial Metabolites in a Coastal Plain Aquifer: The Role of Electron Acceptors" *Environmental Science and Technology*, 29, 2, 458-469.

14. Cozzarelli, I.M., M.J. Baedecker, R.P. Eganhouse, and D. Goerlitz (1994), "The geochemical evolution of low-molecular weight organic acids derived from the degradation of petroleum contaminants in groundwater" *Geochemica et Cosmochimica Acta*, Vol.58, No.2, pp 863-877.

15. Cozzarelli, I.M., R.P. Eganhouse, and M.J. Baedecker (1990), "Transformation of Monoaromatic Hydrocarbons to Organic Acids in Anoxic Groundwater Environment, *Environ. Geol. Water. Sci.*, Vol.16, No.2, 1335-141.

16. Cozzarelli, I.M., R.P. Eganhouse, M.J. Baedecker (1989), "The Fate and Effects of Crude Oil in a Shallow Aquifer II. Evidence of Anaerobic Degradation of Monoaromatic Hydrocarbons" *USGS Water-Resources Investigation Report* 88-4220.

17. Edwards, E., A. and D. Grbic-Galic (1994), "Anaerobic Degradation of Toluene and o-Xylene by a Methanogenic Consortium" *Applied and Environmental Microbiology*, vol.60, no.1,p. 313-32.

18. Eganhouse, R.P., M.J. Baedecker, I.M.Cozzarelli, G.R.Aiken, K.A. Thorn, and T.F. Dorsey (1993), "Crude oil in a shallow sand and gravel aquifer II. Organic Geochemistry" *Applied Geochemistry*, vol.8, pp 551-567.

19. Ehrlich, G.G., D.F. Goerlitz, J.H. Bourell, G.V. Eisen, and E.M. Godsy (1981), "Liquid Chromatographic Procedure for Fermentation Product Analysis in the Identification of Anaerobic Bacteria" *Applied and Environmental Microbiology*, vol.42, no.5, 878-885.

20. Environmental Science and Engineering (1988), "Installation Restoration Program, Phase II, Contamination/Quantification, Stage 2, Volume 1, Tyndall AFB, Florida.

21. Evans, P.J., W. Ling, B. Goldschmidt, E.R. Ritter, and L.Y. Young (1992) "Metabolites Formed During Anaerobic Transformation of Toluene and 0-Xylene and Their Proposed Relationship to the Initial Steps of Toluene Mineralization" *Applied and Environmental Microbiology*, Vol.58, No. 2., 496-501.

22. Geraghty and Miller, Inc. (1991), "Analytical Results Report for Fire Training Areas", Tyndall Air Force Base, Florida..

23. Geraghty and Miller, Inc. (1994), 'Draft Contamination Assessment Report for Active Fire Training Area Site FT-23" Tyndall Air Force Base, Florida..

24. Jackson,A.W., J.H. Pardue, and R. Araujo (1996), "Mineralization in Salt Marshes: Use of Stable Carbon Isotope Ratios, *Environmental Science and Technology*, vol.30, no.4, 1139-1144.

25. Kehew, A.E., and R.N. Passero (1990), "pH and Redox Buffering Mechanisms in a Glacial Drift Aquifer Contaminated by Landfill Leachate" *Ground Water*, vol.28, no.5, 728-737.

26. Kuhn, E.P., J. Zeyer, P. Eicher, and R. Schwarzenbach (1988) "Anaerobic Degradation of Alkylated Benzenes in Denitrifying Laboratory Aquifer Columns" *Applied and Environmental Microbiology*, Vol.54, No. 2., 490-496.

27. Landmeyer, J.E., D.A. Vroblesky, and F.H. Chapelle (1996), "Stable Carbon Isotope Evidence of Biodegradation Zonation in a Shallow Jet-Fuel Contaminated Aquifer", *Environmental Science and Technology*, 1996, vol.3, no.4, 1120-1128.

28. Lovely, D.R., F.H. Chapelle, and J. C. Woodward (1994) "Use of dissolved H_2 Concentrations to Determine Distribution of Microbially Catalyzed Reactions in Anoxic Groundwater" *Environmental Science and Technology*, vol.28, no.7, 1205-1210.

29. Lovely, D.R., and D.J. Lonergan (1990), "Anaerobic Oxidation of Toluene, Phenol, and p-Cresol by the Dissimilatory Iron-Reducing Organism, GS-15, *Applied and Environmental Microbiology*, vol.56, no.6, 1858-1864.

30. McMahon, P.B., D.F. Williams, and J. T. Morris (1990) "Production and Carbon Isotopic Composition of Bacterial CO2 in Deep Coastal Plain Sediments of South Carolina" Ground Water, vol.28, no.5, 693-702

31. McNichol, A.P., E.R.M . Druffel, and C. Lee (1991), "Carbon Cycling in Coastal Sediments: 2. An Investigation of the Sources of SCO_2 to Pore Water Using Carbon Isotopes" in *Organic Substance in Ground Water, Volume 2 Processes and Analytical*, edited by R.A.Baker, Lewis Publishers , Michigan, pp. 249-272.

32. OHM Remediation Services Corp (1994), "Free-Phase Hydrocarbon Delineation and Soil Quality Investigation Fire Training Area 23, Tyndall Air Force Base and (1995) "Pilot Test Work Plan for In-situ Bioremediation Using Microbubbles Active Fire Training Area FT-23" Tyndall Air Force Base, Florida.

33. Seyfried, B., G. Glod, R. Schocher, A. Tschech, and J. Zeyer, (1994), "Initial reactions in the Anaerobic Oxidation of Toluene and m-Xylene by Denitrifying Bacteria" *Applied and Environmental Microbiology*, vol. 60, no.11, 4047-4052.

34. Stumm, W., and J.J. Morgan (1996) *Aquatic Chemistry: Chemical Equilibria and Rates in Natural Waters*, Wiley Interscience, New York.

35. Vroblesky, D.A. and F.H. Chapelle (1994), "Temporal and spatial changes of terminal electron accepting processes in a petroleum hydrocarbon-contaminated aquifer and the significance for contaminant biodegradation" *Water Resources Research*, vol.30, no.5, 1561-1570.

36. Vroblesky, D.A., P.M. Bradley, and F.H. Chapelle (1996), "Influence of Electron Donor on the Minimum Sulfate Concentration Required for Sulfate Reduction in a Petroleum Hydrocarbon-Contaminated Aquifer" *Environmental Science and Technology*, vol.30, no.4, 1377-1381.

37. Wilson, B.H., B.E. Bledsoe, J.M. Armstrong, and J.H. Sammons (1986), "Biological Fate of Hydrocarbons at an Aviation Gasoline Spill Site" *Proceedings of the NWWA/API Conference on Petroleum Hydrocarbons and Organic Chemicals in Ground Water-Prevention, Detection, and Restoration*, National Water Well Association, 78-90.

38. Wilson, B.H., J.T. Wilson, D.H. Kampbell, B.E.Bledsoe, and J.M. Armstrong (1990) "Biotransformation of Monaromatic and Chlorinated Hydrocarbons at an Aviation Gasoline Spill Site" *Geomicrobiology Journal*, vol.8, 225-240.

THE ILLUSION OF CONTROL AND PRECISION ASSOCIATED WITH BASELINE COMPARISONS

David A. Ludwig
Associate Professor
Department of Mathematical Sciences

University of North Carolina at Greensboro
1000 Spring Garden Street
Greensboro, NC 27412-5001

Final Report for:
Summer Faculty Research Program
Armstrong Laboratory

Sponsored by:
Air Force Office of Scientific Research

and

Armstrong Laboratory

July 1996

THE ILLUSION OF CONTROL AND PRECISION ASSOCIATED WITH BASELINE COMPARISONS

David A. Ludwig
Associate Professor
Department of Mathematical Sciences
University of North Carolina at Greensboro

Abstract

Although proper controls are a necessity for valid scientific conclusions, many researchers do not understand how to incorporate proper controls into scientific investigations. Baseline measures are often used as control measures, despite the inadequacies of baseline measures to reflect effects due to the experimental manipulations. The following exposition outlines the problems with baseline observations, and provides an example that demonstrates the increase in experimental and statistical efficiency associated with proper experimental controls.

THE ILLUSION OF CONTROL AND PRECISION ASSOCIATED WITH BASELINE COMPARISONS

David A. Ludwig

Introduction

Most any researcher would agree, that a properly designed experiment must have a control group. However, many researchers do not understand how to incorporate a proper control condition into scientific investigations. And although fundamentally accepted as required, many scientific investigations do not have adequate controls. Researchers often claim a control condition , but in reality, have little more than an observation during the course of the experiment to which they compare the outcome of their experimental manipulation. This, "false control" situation, is commonly seen in experiments in which researchers attempt to use the same subjects for both the control and experimental manipulation. The colloquial term , "own controls" , is often used to describe this type of design. A more technical term for this design manipulation is "crossing". The classical single period cross-over design is one in which subjects receive both the treatment and the control condition. In general, subjects receive both the treatment and control condition in some type of random or counter-balanced random order. The treatment and control condition represent two levels of a single independent manipulation (variable).

The validity of the cross-over design is dependent on two fundamental requirements. First, subjects must experience the exact same manipulations in the control as the experimental treatment condition, except for the experimental treatment which is being evaluated. If the treatment and control conditions differ on other dimensions besides the experimental manipulation, then any comparison between treatments and controls reflects possible differences due to other, "confounding" effects. For example, suppose a new experimental drug is to be tested. Six rats will be crossed over between the new drug and a control condition. The drug is given by injection and then a response variable (dependent) is measured. To insure that the rats had identical manipulations under both

the control and the treatment conditions, rats are given sham injections when in the control condition. This is done to insure that other variables such as handling or injection trauma do not confound the results. If rats in the control condition were left untreated (no sham injection), it would be impossible to separate the effect of handling or injection trauma from the drug effect.

The second requirement is that the order in which the subjects receives the treatments must be, in some way, randomized. This insures that the temporal ordering of the experimental manipulation does not confound the results. If the six rats in the drug experiment were all given the sham injection first and then crossed over into the drug condition, there would be no way to separate the drug effect from the effect of the treatment order. Differences between drug and sham treatments might then be a function of maturation, acclimatization, seasonal variation , or any number of other confounding factors. When an experiment is "confounded in time", there will be many reasons, beyond the experimental manipulation, for an observed effect.

<u>The Baseline S.N.A.F.U.</u>

It is not uncommon, both in the literature and at scientific meetings, to encounter the "baseline control study". Subjects are measured pre-treatment, the treatment is applied, and then they are measured post-treatment. Researchers then claim, with this manipulation, subjects served as their own controls. They then proceed with a paired t-test and conclude that the treatment was effective. Experimenters are in denial if they believe they have a true control condition. What they have is a "false control", in which the effects of the treatment cannot be separated from the multitude of confounding factors which have an effect during the passage of time. Their so called control condition, occurs temporally before the treatment. Therefore, any difference seen between pre and post-treatment measures can be attributed not only to the treatment, but to any number of effects carried by the time confound.

This type of experimental manipulation has historically been referred to as the "one-shot study". Others have used the terminology "pseudo design" or "pseudo control". Where pseudo, according to Webster, means false, apparent, or erroneous. Although this type of experimental manipulation may indicate a difference between pre and post-treatment measures, there is no way to attribute this difference to the treatment that was applied. Why? Because there is no true control. These pre-treatment control condition designs, lack both fundamental elements for valid cross-over manipulations. Since the control condition is measured before any type of experimental manipulation, things that are done to the subjects, other than the treatment, are not represented in the control measures. The effects of testing, knowledge of the experimental setting, and familiarity with the experimenter are things which are not represented under the control condition. Thus, the control condition is different from the post-treatment condition on any number of variables besides the treatment being investigated. No amount of hand-waving, excuse making, or statistical manipulation can remedy this problem. Until subjects are properly crossed-over between the treatment condition and a true control condition, can the researcher attribute the difference to the experimental manipulation.

Proof by Example

My own experiences as a consulting statistician provide me with an number of examples which I could use to prove the point, that baseline controls are often misleading and highly bias. Recently, I was involved in the design of a head down tilt study in which four days of head down tilt (HDT) were to be compared to four days of upright control (UC). The researchers were investigating a number of physiological changes, one of which was resting heart rate. The study was to be conducted on six healthy rhesus monkeys. At the initial design meeting, the experimental protocol was discussed, and the researchers proposed measuring each monkey prior to HDT and using this as a baseline control. Then, after four days of HDT, a second measure would be taken and compared to the pre HDT baseline as evidence of an HDT effect. After informing the researchers of all the problems associated with such a design, the protocol was changed to a proper

cross-over manipulation, with a true control. For this design, half of the monkeys would receive HDT followed by UC and half UC followed by HDT. A between treatment interval of two weeks was used to allow for physiological resetting during the cross-over period. When monkeys were in the UC condition, they were handled, feed, and tested in exactly the same manner as the HDT condition. The only thing that was variable between the two treatment conditions was position, which was maintained at 60 degrees head-up in the UC condition. Thus the UC condition provided a true control to which HDT could be compared.

Although a baseline measure, taken prior to treatment, would have no bearing on the comparison of HDT to UC , one of the researchers felt uncomfortable without some type of baseline measure. So a baseline measure was taken before any type of experimental manipulation. This measure would have served as the control in the original proposal, but given the true control of the UC condition, provides little more than descriptive information about the condition of the monkeys before entering the experiment. Although this information may be useful in detecting monkeys who are ill or who may be outside some reasonable standardized value, it is for the most part worthless when investigating the effect of HDT. It does however, provide me with the necessary data to analyze this experiment as if it were run under the original protocol (baseline/post-treatment) and compare the results and conclusions to the results and conclusions of a properly run cross-over design with a true control. The data presented is actual data and has not been augmented or manipulated in any way.

<u>Results and Conclusions (False Controls)</u>

Table 1 gives the raw data, descriptive statistics, and statistical test results for the HDT condition. The results show a decline of 12.82 bpm from the baseline measure. If a paired t-test were run on this data, the researcher would claim a "statistically significant" difference and conclude that the decline in heart rate was due to HDT. There is no denying that a decline was observed. The problem is that there is no way to attribute the

observed decline to HDT. Perhaps it was the effect of having to restrain the monkeys, or maybe it was the stress induced by the myriad of tests and blood samples that were collected over the four day period. There are an infinite number of reasons for the decline over this four day period, many of which the experimenter is not even aware of. Perhaps monkeys were responding to the handling that was required to obtaining all the scheduled tests. The fact is, all of these explanations are as plausible as the HDT hypothesis. Good experimental design does not require the researcher to debate competing reasons for an observed difference. Who's to say what did and did not contribute to the observed difference.

TABLE 1

PRE AND POST HDT DATA FOR HEART RATE (BPM)

Monkey	Pre HDT	Post HDT	Difference
1	161.7	151.8	-9.9
2	152.8	148.8	-4.0
3	116.3	109.2	-7.1
4	154.9	133.3	-21.6
5	165.2	135.3	-29.9
6	127.0	122.6	-4.4

Mean Difference = -12.82
$SD_{difference} = 10.57$
$SE_{difference} = 4.32$
$| t |_{df=5} = 2.97$
$P < .05$

Before proceeding to the UC condition, a word or two concerning the role of statistical tests might be helpful. In a well planned design, there is only two reasons for the observed difference. One is the experimental manipulation (e.g., HDT) and the other is sampling variation associated with the randomization of the subjects to the experimental

conditions (i.e., chance model). If the results of statistical tests indicate that the observed difference is up and above what might be expected solely as a result of sampling error, then in a properly designed experiment, the treatment effect remains as the only other logical reason for the observed difference. On the other hand, if the design is flawed, and there are many reasons other than sampling variation for the observed difference (confounds), the results of statistical tests are moot. The low P value associated with the statistical test may have "psychological value" for some, but it is no help in determining if there is a treatment effect. Unfortunately, the results sections of many published articles are littered with meaningless P values in an effort to give credibility to a flawed experiment. In reality, they make the situation worse by suggesting that these tests give credibility to the research hypothesis. The results of the paired t-test between baseline and day 4 observations indicate that it is unlikely that a difference of 12.82 bpm would have been observed if there is no effect of HDT ($P<.05$, Table 1). Given this unlikely result, what can be concluded? That there was an effect of HDT? Hardly! Although the chance model has to some degree been discounted as the reason for the observed 12.82 bpm difference, there are numerous other alternative explanations for this difference. These other alternative explanations exist because of the design flaws inherent to the baseline/post treatment protocol.

Baseline to Day Four Differences During Upright Control

The UC condition produced virtually the same decline in heart rate from baseline to day 4 as the HDT condition (12.57 bpm, Table 2). The difference between the HDT and UC differences is less than one bpm and would be considered "non significant" by any statistical, clinical, or scientific standard. At this point, some would be tempted to perform another paired t-test on the UC data and compare the results to the paired t-test conducted on the HDT data (Table 1). This would be incorrect. Comparison of separate paired t-tests between baseline and post treatment observations conducted within the two treatment conditions does not test that HDT and UC differences are different. What needs to be tested is the difference between the two difference means (i.e., (-12.82)

-(-12.57) = -.25). The test that -.25 is not equal to zero requires that this difference be compared to the appropriate standard error of the difference between the two difference means. The appropriate standard error is not considered when a simple comparison is made between the results of within treatment condition t-tests.

TABLE 2

PRE AND POST UC DATA FOR HEART RATE (BPM)

Monkey	Pre UC	Post UC	Difference
1	153.0	157.0	+4.0
2	151.1	149.8	-1.3
3	135.7	124.3	-11.4
4	166.8	131.8	-35.0
5	171.8	147.1	-24.7
6	146.2	139.2	-7.0

Mean Difference = -12.57
$SD_{difference}$ = 14.73

Since each monkey received both the HDT and UC condition, differences generated between the baseline and day 4 measures for each subject, can be compared with a paired t-test across the two experimental conditions (Table 3) This test is mathematically equivalent to the test of interaction between treatments and time. Tests of interaction test if the effect of one variable is the same at each level of a second variable (i.e., Is the effect over time (baseline to day 4) the same for each treatment condition (HDT versus UC).). By definition, interactions test for differences between differences. The results of this statistical test indicate that the difference between these two differences was not differentiable from what would be expected given only sampling variation ($t(5)=.07$, $P>.50$). Thus, there is no compelling evidence for an HDT effect, since the decline in

heart rate in both the HDT condition and the UC condition was the same. This is in conflict with the conclusions that were reached when the pre treatment (baseline) measure were used as the control. Hum!

TABLE 3

PRE TO POST TREATMENT COMPARISON BETWEEN HDT AND UC FOR HEART RATE (BPM)

Monkey	HDT Diff.	UC Diff.	Difference
1	-9.9	+4.0	-13.9
2	-4.0	-1.3	-2.7
3	-7.1	-11.4	+4.3
4	-21.6	-35.0	+13.4
5	-29.9	-24.7	-5.2
6	-4.4	-7.0	+2.6

Mean Difference = -0.25
$SD_{difference} = 9.29$
$SE_{difference} = 3.79$
$|t|_{df=5} = 0.7$
$P > .50$

When compared to a control situation, it is evident that the observed change from baseline to day 4 cannot be attributed to HDT, since the control condition demonstrated the same decline. The decline over the 4 day period would seem to be a function of how the experimental material was handled. There is no evidence that the HDT treatment attenuated or increased this decline. At this point, there is no evidence for an HDT effect. Fatigue, boredom, the taking of blood samples, test batteries administered during the four day protocol, or experimenter/monkey interaction are just a few of the many possible reasons for the decline from baseline.

The Problem With Gain Scores

Gain scores (baseline to treatment differences) are notoriously unreliable. The problem is that measurement error associated with baseline and treatment measures is compounded when a difference score is calculated. I don't want to get into a statistical or psychometric discussion of the hazards of using gain scores, since numerous references are already available on the subject (Rogosa & Willett, 1983). Suffice it to say, that the observed variance of a gain scores (differences) will always be more than the original measures that are differenced, unless the original measures are perfectly reliable (not likely). Since reliability is defined as the ratio of true score variance to observed variance, increases in observed variance will result in reduced reliability. This decrease in reliability is a function of increased observed variance resulting from the differencing. More measurement error, means higher experimental error (greater variance), which results in a reduction in statistical power. As discussed above, when comparing differences between differences, the proper error components must be considered when performing statistical tests. Separate, within treatment condition comparisons, do not consider the increased variation associated with differencing.

Results and Conclusions (True Controls)

Baseline information is not required when a true control condition is available. In fact, the baseline measure is somewhat of a distraction, and as stated above, creates unwanted error variance. A comparison of the HDT to UC condition at day 4 is all that is required. Since the HDT/UC order was counterbalanced, there is no time confound associated with the HDT/UC comparison, nor is there any bias associated with where the subjects started before they began the experiment. The true cross-over design, in essence, randomizes the starting point of the subjects so there is no need to reference measures from values obtained before the treatment was applied.

The average heart rate at day 4 was 133.5 bpm for the HDT condition and 141.5 for the UC condition. The results of a paired t-test indicates that this 8 bpm difference is up and

above what might be attributed to sampling variability (Table 4). Given that the design was well controlled, with no confounds due to the order in which the treatment conditions were applied, there is only one other alternative explanation for this 8 bpm difference. HDT! Although the true reason for the observed effect is never actually known, unlike the baseline/post HDT comparison, this 8 bpm difference can be logically attributed to the effect of HDT.

TABLE 4

COMPARISON BETWEEN THE HDT AND UC TREATMENT CONDITIONS FOR HEART RATE (BPM)

Monkey	HDT	UC	Difference
1	151.8	157.0	-5.2
2	148.8	149.8	-1.0
3	109.2	124.3	-15.1
4	133.3	131.8	+1.5
5	135.3	147.1	-11.8
6	122.6	139.2	-16.6

Mean Difference = -8.0
$SD_{difference} = 7.56$
$SE_{difference} = 3.09$
$|t|_{df=5} = 2.60$
$P < .05$

Although the HDT effect was now smaller than that previously estimated (8 bpm versus 12.82 bpm), this difference was still differentiable from sampling error . Since no differencing from baseline measures was involved in determining the 8 bpm estimate of treatment effect, the standard error associated with this estimate was less than the estimate involving baseline measures (3.09 versus 3.79). This illustrates, the inflation of error variance due to differencing (see Table 4).

Proper Scientific Behavior

A common response to the hazards of pre-post data is, "But this is all I have or all I can get!" The fact that a researcher cannot perform a proper experiment, for what ever reason, does not change the fact that the conclusions from such investigations will always be suspect. The problem is not in the data or design per say, the problem is what the researcher wants to say about the results. Rather than admitting that the change from pre to post treatment cannot be directly attributed to the treatment intervention, gross statement of cause and effect are often advanced. Adding insult to injure, investigators perform a variety of meaningless statistical hypothesis tests with the idea that such procedures, if successful (i.e., P values less than .05), absolve them and the data from any problems associated with the experimental design. The idea that all is forgiven, as long as P is less than .05, is to say the least, naive. There is nothing scientific about the incorrect application and interpretation of statistical tests. Science would be better served if investigators and journal editors would adamantly resist the notion of statistical tests in pre-post investigations.

There is no excuse for incorrect design, especially when it would have been possible to correctly manipulate the experimental material (i.e., design errors created by the researcher). Yet, there is nothing inherently "evil" with pre-post data. Pre-post data often results from what are commonly referred to as observational studies. These studies do not permit the same level of control as might be available in a laboratory setting. Researchers must understand, that the inability to correctly manipulate subjects in an experiment, restricts what can be said about what caused the results. Observational studies often produce useful scientific information and should not be necessarily rejected out of hand. They should be rejected when investigators attempt to attribute the results to single cause and or support their ideas with statistical tests.

Summary

Proper experimental design can eliminate the need for baseline referencing while preserving the proper, unbiased comparisons, between treatment conditions. Uncovering the HDT effect required correct design coupled with correct statistical manipulation. The inclusion of baseline measures serve only to complicate the issue. Baseline measures do not possess control information and when differenced from treatment values, inflate the experimental error variance. They are merle observations taken before the experimental material is subjected to the experimental manipulation. Since baseline measures are collected prior to experimental manipulation, they contain no information with regard to the effect the experimental process had on the experimental material. True control measures provide a comparison between the combined effects of experimental manipulation and treatment, and experimental manipulation alone. This comparison is at the heart of the matter (no pun intended). It answers the question, "Is the observed effect a result of the treatment or the experimental manipulation?". Or, "Is the observed effect in the treatment group up and above that attributable to only the experimental manipulation?". Since baseline measures occur prior to the experimental manipulation, these questions cannot be answered with a baseline/post-treatment design.

Unfortunately, researchers harbor a persistent "cognitive illusion" that a pre-treatment baseline measure is a necessity, reflecting good scientific practice. In reality, baseline measures should initiate a red flag. More often than not, baseline measures are associated with poor experimental design or observational studies in which manipulation of the experimental material is not possible. Hopefully, after reading this exposition, researchers and consumers of research will better understand what can and cannot be concluded from baseline/post treatment studies and use caution when evaluating results from such investigations.

References

Rogosa, D. R., & Willett, J. B. (1983). Demonstrating the reliability of the difference score in the measurement of change. *Journal of Educational Measurement*, 20, 335-343.

DESIGNING INSTRUCTION FOR DISTANCE LEARNING

Robert G. Main
Professor
Department of Communication Design

California State University, Chico
First and Normal
Chico, CA 95929-0504

Final Report for:
Summer Faculty Research Program
Armstrong Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, DC

and

Armstrong Laboratory

August, 1996

DESIGNING INSTRUCTION FOR DISTANCE LEARNING

Robert G. Main, Ph.D.
Professor
Department of Communication Design
California State University, Chico

Abstract

This study explores new models and methods for use by instructional designers and developers for exploiting the technology of distance learning in creating more effective instruction. Organizations are beginning to change their question from "Should we be doing distance learning?" to "When are we going to begin distance learning?" Traditional classroom models for instructional design and development need modification for application in a distance learning environment. An extensive review of the literature was conducted to provide both practical guidelines and theoretical considerations in transforming traditional instruction to distance delivery lessons. The study looks at many aspects of the design process from selection of instructional strategies and activities to the role of the instructor in the design and delivery of the content. Particular attention is paid to classroom interaction and learner participation, collaborative learning and contingency planning. An integrated behaviorist/constructivist model for distance learning programs is discussed along with learner autonomy and locus of control. Media presentation is also examined for a variety of transmission systems.

DESIGNING INSTRUCTION FOR DISTANCE LEARNING

Robert G. Main, Ph.D.

Traditionally the effectiveness of distance learning instruction has been measured against the learning outcomes of a "regular" class using a common test instrument. Similarly, the cost effectiveness of distance learning technology is to compare it with the cost of the traditional classroom programs. A study by North Central Regional Educational Laboratory concluded, "Effectiveness is not a function of the technology, but rather the learning environment and the capability to do things one could not do otherwise. Technology in support of outmoded educational [practice] is counterproductive,,,, Technology works because it empowers new solutions (Jones, et al, 1995, p. 6)." This study explores new models and methods for use by instructional designers and developers for exploiting the technology of distance learning for creating more effective instruction.

The rapid rise of digital telecommunication and the transformation of media from analog to digital formats have opened the door to instructional delivery possibilities that have never been seen before. Archives of lunar travel and space exploration, video and audio files of news and information around the world, and entire library reference services are being made accessible to learners of all ages through the Internet. These resources and technologies are no longer restricted to an elite community and they are changing the nature of education and training forever as a result (Capell, 1995). Organizations are beginning to change their question from "Should we be doing distance learning?" to "When are we going to begin distance learning?" The very nature of learning is changing with the new opportunities provided. The demand will be on greater access, improved interfaces and more interesting and stimulating presentations by an expanded learner base. Satellite presentation at present is the dominant delivery technology and will likely remain competitive for a considerable period of time. The change to network-based multimedia delivery systems is inevitable, however, as the initial capitalization costs for fiber optics and other wide band installations are amortized through user volume.

Defining Distance Learning

Distance learning terms are defined by a national task force of distance learning scholars chaired by the American Council on Education as follows:"

> *Distance learning* is a system and a process that connects learners with distributed learning resources. While distance learning takes a wide variety of forms, all distance learning is characterized by:
>
> > Separation of place and/or time between instructor and learner, between learners and/or between learners and learning resources;
> >
> > Interaction between the learner and the instructor, among learners; and/or
> >
> > between learners and learning resources conducted through one or more media; use of electronic media is not necessarily required.
>
> The *learner* is an individual or group that seeks a learning experience offered by a provider.
>
> The *provider* is the organization that creates and facilitates the learning opportunity. The provider approves and monitors the quality of the learning experience. Providers include schools, colleges and universities, businesses, professional organizations, labor unions, government agencies, libraries, and other public organizations (Guiding Principles for Distance Learning in a Learning Society, 1996.)

This definition would include correspondence courses by mail as well as instruction via the Internet, by television or radio broadcast, individualized computer programs, video and audio tape, CD-ROM, and full motion interactive audio/video and data in real time. Each form represents particular challenges and opportunities to the instructional designer. For illustration, this paper focuses on real time interactive audio/video systems. The design principles, however, should apply to all forms of distance learning.

The Delivery System Is the First Consideration

Most Instructional Design models do not consider delivery modes until the Analysis Phase is complete. Usually they are selected after the needs assessment has been completed, the learning objectives developed with their criteria for measurement, the learning strategy selected and instructional activities sequenced. The assumption is that the learning will occur in a traditional classroom. This is entirely appropriate as 95% of formal education and training is still conducted with students assembled in a room with an instructor delivering or at least controlling the instruction in a real time, face-to-face interaction. As communication technology capabilities improve and education and training requirements expand, the need for alternative delivery methods is increasing.

Instructional strategies and activities for distance learning involve all the components of instructional design with the added complexity of distance delivery (Wagner, 1990).
In his *Report on Distance Learning Technologies,* Capell (1995) states:

> The landscape of educational decision making has changed with advances in instructional technology. If the ISD model (or other variation) is assumed as the basis for course development, then not only can we say that the decisions themselves have changed, but the timing as to when the decisions are made has also changed in reference to this model....[T]his means that it is now appropriate to ask... [about the] use of technology at the same time that instructional goals are considered as opposed to waiting until we reach *develop instructional strategy* or *develop and select instructional materials.* This is a big change over past practice, in which the use of instructional technology would have been considered only well after the objectives of the course were determined (p. 50).

Effective Learning

There is a dramatic shift occurring among educators in the definition of learning from the traditional transfer of knowledge and skills model which implies an expert presenting information to an engaged learning model with all the implications for responsibility, control and interactions. While this phenomenon is not unique to distance learning, the application of the fundamental principles require greater attention when the instructor and students are not face to face in a classroom. The variables for selecting the delivery strategies and designing the learning activities for engaged learning are (Jones, et al, 1995, p. 7):

-learners engage in authentic and multidisciplinary tasks
-assessments are based on performance of real tasks
-learners participate in interactive modes of instruction
-learners work collaboratively
-learners are grouped heterogeneously
-the instructor is a facilitator in learning
-learning is by exploration

Analysis of the Delivery System

The first step in designing instruction for distance learning is the analysis of the delivery system. The assumption is that the instructional designer will not have the luxury of specifying the delivery system, but will have to work with the capabilities that exist. It is also axiomatic that the lesson must be developed for delivery over the least sophisticated technology available to its intended users (Bradley and Peacock,1996). If some distance learning users are restricted to transmission rate of 19.2 kbs, for example, that becomes the limiting factor in the instructional delivery. This is the weakest link rule.

Selecting the Instructional Strategies

There are four overall considerations in creating an engaged learning environment in a distance learning course:

The learners must assume greater responsibility for their own learning. They must exert initiative and greater self-regulation than students in a traditional classroom setting.

They must be more aware of their knowledge and skill acquisition. Self assessment and knowing *how* to learn is as important as *what* they learn. Constructing effective mental models of the subject domain is critical for knowledge transfer and connection.

The learners must be motivated to learn. They derive excitement and pleasure from learning that energizes them to take additional steps to refine their knowledge and problem solving skills.

Teamwork is emphasized. The learning is collaborative to instill the value of other's viewpoints and the ability to work with them skillfully.

Designing Learning Activities

Learning activities should be relevant, challenging and authentic. The knowledge and skills must be explicit to the learner's self-interest--not the organization, although that benefit should be implicit. The tasks assigned must be sufficiently difficult to be mentally or physically interesting, but not to the point of sustained frustration. The learning activities are authentic when they replicate behaviors beyond the classroom setting in real-life tasks. The activities should be in a context that links part tasks to complex tasks. In other words, students should learn by doing whenever possible.

Instructor Role

Activities should be learner centered instead of instructor centered. In a distance learning environment where interruptions in communications may occur unexpectedly and frequently it makes sense to move the locus of instructional activity to the student as much as possible. It is also pedagogically sound to make the instructor more a facilitator or guide than the presenter of the lesson content. A well designed distance learning lesson can provide a rich learning environment by creating opportunities for students to work collaboratively, to solve problems, conduct research, do authentic tasks and simulations, and share knowledge and responsibility. This often requires a different set of skills for the instructor as mediator, model and coach. Instructors must constantly monitor and adjust to student needs for information, resources and problem solving strategies. This can be a difficult transition for instructors accustomed to being the center of knowledge and attention.

Learning Assessment

Learning assessments should allow students to demonstrate their knowledge and skills in authentic tasks or projects. Performance based assessments should involve planning and execution as well as self and peer evaluations of products, presentations and debriefings. In team projects, involving the group in the

development of the assessment measures and procedures is itself a challenging and meaningful learning activity. The ideal situation is where the instructional activities and the assessment are seamless and transparent to the learners as they move from one to the other. It is critical that standards are well defined and equitable.

Grouping Learners

Collaborative learning activities often involve small groups or teams of two or more students within proximity or at different learning sites. Although each learner's role and tasks may be different, all members collaborate to accomplish a joint objective or product. Assigning students in proximity with each other is desirable when all other factors are equal. Students can work together off-line with the rich interaction of face to face interaction and without the telecommunication costs. There may be other factors, however, where it is desirable or necessary to create groups with members from geographically separated sites. It may be important, for example, that students have experience in distance collaboration using the telecommunication technology. Exercising the technology may be a learning objective for the class.

A technique used at California State University, Chico for a course taught over the Internet, involved the pairing of students with varying prior experience. A required class for Communication majors, the students beginning knowledge ranged from extensive technical skills to never having used a computer. A skills assessment administered during the first class session provided a rank order of class members for skill level. The students were then assigned as two person teams with the number one ranked student and the lowest ranked student in one team, the second ranked student and the next to lowest ranked student in another team. The process was repeated until all students had been paired. Students were provided an extensive syllabus for the course explaining the learning objectives for the class and a detailed description of how the class would be taught using the Internet. The purpose of the teams was explained and that learners of varying abilities had been purposely assigned to provide peer tutoring for less experienced students.

The Team Approach

The instructional design/development team should be assembled with the customary instructional designer and one or more subject matter experts supported by the necessary media production expertise to create the lesson materials. In addition, the team should have the services of a telecommunication technician either as a member or as a readily available consultant.

It would be ideal, of course, if the instructional designer could analyze the instructional needs and appropriate strategies and then design the delivery system that best accommodates the learning events developed to achieve the desired learning outcomes. Practically speaking, the instructional designer/developer is given the task to produce a course to be presented over an existing system. It is imperative, therefore, that the designer have explicit knowledge of the capabilities and limitations of the system and how it operates. If possible, the designer should have taught or taken a class using the distance learning system.

Instructor Involvement in the Instructional Design

It is a good strategy to involve the instructor in the design process for the distance learning class. Few distance learning classes are "new" courses. In an examination of 81 distance learning courses offered over a four period at California State University, Chico, Dixon (1996) found only five had not been offered previously as traditional classroom presentations. Furthermore, the five that had not been previously taught were specially tailored workshops designed for one time scheduling to meet a particular need for adult education. Although this is a university where curriculum changes are not as common as industry training programs, the development of a totally new class for distance learning delivery is most likely the exception for corporate training needs as well.

It is only logical that the best candidate to teach the distance learning class is the instructor who has been teaching it in the traditional classroom. The instructor serves not only as a subject matter expert, but also as the expert on pedagogy for the class. His or her experience is invaluable in the restructuring of the learning strategies and activities to fit the limitations and capabilities of the distance learning delivery system

Convert Easiest Courses First

If the situation permits, the introduction of distance learning classes should start with those courses most easily converted. This usually means those where the mode of instruction is primarily lecture based. The planning team can go through the course listings currently being offered and rate the classes as to the level of effort required for conversion (e.g. minimal, moderate, extensive). Not only does this allow for an easier transition for the design and development team, it also provides a lower slope to the instructor's learning curve for distance learning presentations. An obvious bonus benefit for including the instructor in the design and development process, is the knowledge gained of the delivery system. By beginning simply, the confidence of the instructor in the system is reinforced by success.

The design and development team will also learn from their experiences. With the simpler conversions as a starting point, the complexity of the curriculum development is incremental so that the level of effort for each course preparation remains relatively constant.

Formative Evaluation

Formative evaluation is critical for all instructional design. Braden (1996) has incorporated it as an explicit function of each step in his adaptation of the Dick and Carey ISD model. Main (1993) shows validation and feedback as an integrated activity for each phase of his motivation integrated ISD model. The importance of validating each activity in the design process is elevated in distance learning development. In traditional classroom instruction, the rich and immediate feedback from the students permits the instructor to make changes in the delivery on the fly. Good instructors are continuously monitoring their classrooms for visual and audio cues that indicate students are attentive and actively engaged in the learning process. Corrections can be made individually and collectively to modify the planned activity to insure student comprehension and involvement. This is not possible in distance learning environments. Even in the most sophisticated virtual classroom, the technology degrades the quality of interaction. In most distant learning systems, the students cannot be seen by the instructor or can only be seen when they wish to speak. The small screen and low fidelity of wide area classroom views makes it difficult to distinguish individuals let alone their facial expressions. Pilot testing of every aspect of the lesson is essential. Not only is feedback limited in distance learning environments, but the flexibility of the instructor to change delivery is restricted by the technology and time pressures. It is much more difficult for students to hang around and ask questions of the instructor after class unless this provision has been built into the lesson during its design so that network time is available

Learner Motivation

Main (1992) found technology required greater attention to learner motivation in the design of instruction. In traditional classroom presentation, the instructor is largely responsible for attracting and maintaining learner attention. Personal anecdotes or examples from the instructor's repertoire of experience can be inserted to establish relevance for the learner of the knowledge and skills being taught. The level of difficulty of the lesson can be adjusted on the fly to bolster the learner's confidence. Encouragement and feedback regarding learner performance is immediate and contextually rich. Good instructors do these things automatically. They are in continuous rapport with the performance and mood of the class members.

Where interaction is intuitive in the traditional classroom, it must be carefully planned for distance learning environments. Bradley and Peacock (1996) express concern that distance education may not allow for that vital human contact with instructors, resource people and other students that is such an essential part of a good education. The challenge for the instructional designer is whether distance learning will be just a poor substitute for more personal (and arguably more effective and desirable) traditional means of teaching, or whether it can be used for a qualitatively different type of instruction.

A great deal has been written of the ability of computer delivered instruction to individualize instruction to the learner's needs. However, the most advanced artificial intelligence technology cannot begin to match the ability of even a mediocre instructor to respond to the dynamics of a classroom. It is imperative, therefore, that when instruction is to be mediated by technology, the greatest attention must be paid to designing learner motivation into the presentation.

Classroom Interaction

The successful expansion of distance learning as an alternative to the traditional classroom is dependent upon the instructional design to approximate the richness of the interaction that occurs face-to-face(Main & Riise, 1994). There are six factors which should be considered in designing distance learning interactions. They are: 1) The amount (frequency and length of dialog; 2) Type (instructor-student, student-student, and student-course content); 3) Timeliness (a continuum ranging from full duplex conversation to asynchronous exchanges with days of delay); 4) the Method of interaction (refers to the medium and channel used from voice to text to non-verbal gestures); Spontaneity (refers to whether the transactions are preplanned or ad hoc exchanges triggered during the presentation); and Quality of the interaction (intensity or emotional involvement, relevance, depth, formality, and opportunity).

Interaction always occurs within a context. There are numerous factors that may be affected by, or have an effect on, interaction in distance learning. These factors can generally be classified as those concerned with the course and those concerned with its delivery, i.e., the communication technology. Course or curriculum variables include the subject matter, student characteristics, instructional strategies and activities, media used, and instructor attributes. Variables associated with the delivery of the instruction are concerned with the transmission capabilities of the network (bandwith and data rate) and hardware and software configurations of the origination and distance learning sites. Class size is an overarching variable in instructional interactions. As the size of the class increases, the chance of interacting with the instructor dwindles no matter how sophisticated the communication technology or elegant the instructional design.

Distance learning may depend even more on instructor charisma and style than the traditional classroom which means instructor characteristics are important in their effect on interaction. There is a large body of literature available on instructional process, but despite the scrutiny of what goes on in the classroom, teaching remains very much an art form. A study by Fulford and Zhang (1993) suggests the perception of overall interaction is a greater predictor of student satisfaction than actual personal interaction. In their study of a class of 123 students in five locations the perception of overall interaction (self-report) had a strong correlation with learner satisfaction despite the number of personal interactions. This "vicarious" interaction effect should not be too surprising. The appeal of game shows and talk shows is largely the interaction between host and guests or contestants and their success is dependent on the artistry of the host in generalizing audience identification with him or herself and the topic.

While classroom interaction is almost universally considered an enrichment to the learning process, there is some evidence that it is not a critical component for learning. Studies by the Navy of video televised training (VTT) instruction found a significant reduction in interaction in VTT classes when compared with traditional classroom presentations. However, learning outcomes measured by the same multiple choice exam were identical for the groups (Wetzel, 1996). This is not unusual in the literature. Study after study indicates student achievement in distance learning classes is equivalent or superior to traditional classroom student achievement (Salomon & Clark, 1977; Ritchie & Newby, 1989). In a meta

analysis of media use in instruction, Simpson (1993) found, "Achievement is similar to conventional education with interactive television or video teletraining, and with correspondence 'telecourses'"(p 153). Most distance learning studies are flawed, however, in their inability to control contaminant variables through random assignment to treatment and control groups or the use of matched pairs. They are generally case studies conducted in the field with the possibility of many differences in demographics between traditional classroom students and those taking the instruction at a distant location. At the very least, the distant learner by definition is being offered an opportunity for learning that might not otherwise be available without considerably greater effort and expense on the part of the learner. This tends to create a student who is more appreciative of the opportunity for the learning experience and consequently more dedicated.

Lacking solid contrary evidence, it is only commonsensical to maximize interaction opportunities in the lesson design.

Learner Participation

There is obviously an overlap in the concepts of classroom participation and classroom interaction as discussed above. For purposes of distinction, participation is defined as learner involvement in the instructional process. Participation is a more generic term that subsumes classroom interaction. It can be broadly categorized into classroom interactions (student-teacher, student-student and teacher-student), group interactions, (projects, problem solving, team drills), interactions with learning materials and resources (research reports, reading assignments, homework activities), intellectual interactions (critical thinking and higher order cognitive skills such as analysis, synthesis, evaluation), and emotional involvement (attitude, attachment, motivation). Emotional involvement is more properly addressed as a function of mode and method of presentation.

The criticism of most distance learning systems in the past has been the imbalance between the amount of time spent by experts presenting information and the arrangements made for the learner to interact with the content, with the instructor and with other learners. This criticism is also valid for classroom presentations where an instructor (expert) lectures to the students. This large lecture model is popular with instructors (after all, they are in the position of authority and control) and administrators because of its low cost per unit. It persists despite the mounting body of evidence that learner centered strategies are more effective . But students do learn. It is superior to a textbook and to recorded lecturers in that there is some spontaneity and ability to adjust to student feedback. It is used widely in "educational television" programming where courses are presented by television station broadcast or satellite transmissions to large audience groups. The participation is analogous to talk radio or television where the limited audience interactions are presumed to be representative of the wider audience.

In the lecture mode, student participation with the content can be accomplished with assignments performed outside the class period. These may be reading assignments, research papers, or problems to solve. Students may be required to watch a film or video or listen to an audio tape. Group activities and collaborative learning projects may be assigned that will require students to interact with each other outside the regular class period and report or demonstrate their work to class at large.

Asynchronous Interactions

There are a number of distance learning systems in which synchronous interactions are not feasible. The oldest, of course, are the correspondence courses that have been offered for more than a century and are still attracting students. They are primarily print-based although there are audio and video tape versions as well. The Public Broadcasting Service (PBS) member stations offer a number of telecourses, often in conjunction with local colleges and universities, that are received by students on their television receivers at home or office. If college credits are to be awarded, registration, graded assignments and tests are usually administered by mail or at scheduled meeting times in a local classroom. Professors receive a stipend for providing advising, grading and exam proctoring.

The latest in asynchronous instructional interactions are the on-line programs springing up across the nation using the Internet, commercial on-line services or corporate data networks. The delivery of instruction synchronously changes the nature of the teaching function from lecturing to coaching (Mason and Aye, 1989). The challenge of communicating without visual or audio cues, coupled with the lack of immediate feedback can lead to anxiety or misinterpretation of the intended message which makes faculty feedback particularly important. Instructors need special training in the nuances of on-line communication. Consider the reliance placed upon body language and facial expressions in traditional classroom instruction. The instructor may respond to student comments with a smile or a nod and this non-verbal communication is meaningful and satisfactory to both parties. Failure to respond to questions on-line can be viewed as rejection by the communicator (Hedegaard, 1996). The use of sarcasm, humor and irony must be carefully composed or avoided. They can be easily misinterpreted when delivered in writing because this is a very literal environment.

There is a positive side to the on-line computer-mediated instruction. According to Mason and Kaye (in Harasim, 1990), the lack of visual cues creates a unique democratic atmosphere. On-line class groups can work together, dialog, debate, and converse indefinitely without being prejudiced by race, gender, appearance or even personal charisma. Individual contributions are valued on their merit and content of the message is the primary focus. This is an ideal situation for developing critical thinking tools and creative problem solving techniques. Of course descriptive identifications of participants can be provided when necessary.

Learner Autonomy

Learner autonomy is participation by the learner in the determination of their learning outcomes. In various forms it is referred to as constructivism, student centered learning, locus of control, et al. It begs the larger question, "Who knows best what behaviors the learner must master?" While there is general agreement that subject matter experts and instructional designers/developers are best prepared to package and deliver the instruction, there is less agreement that the learner is in the best position to determine what knowledge and skills they need to acquire (or even when they know what is best for them that given an opportunity to choose they will select what is best).

It is certainly true that self selection will provide a more motivated learner (a powerful argument for just-in-time training). .. It seems less self evident that learners are in the best position to determine their own needs, or, in the case of corporate training, whether their needs meet the expectations and needs of the organization. Tasks and employee performance levels seem best determined by experienced practitioners and managers. Scholars in the field may have an even broader view of knowledge and skills needed beyond the narrow context and situation of a specific job in a particular company.

Even if the performance desired could be explained in detail under the conditions of performance and the level of proficiency required, it is still possible that learners would lack the self-assessment ability to make correct curricular decisions. Students are notoriously inaccurate in evaluating their own level of knowledge and skill levels. A study by Niculescu-Maier (1995) found student expectations of their grade varied on average one full grade level from the grade awarded by the professor. The range was as much as three grade levels above the actual grade to two grade levels below. Furthermore, it is human nature to wish to increase our knowledge in those subject areas we like (and are often most proficient) and avoid those subjects we dislike. Main (1986) found the most common reason listed by students in a beginning communication class as to why they had chosen a communication major was that no math was required.

Behaviorist v. Constructivist Compromise: Perhaps some accommodation can be reached between central design of curriculum and design to student need if we accept the optimal learning curve model for behavioral v. constructivist design methods. A learning curve is constructed of the knowledge, skill or affect to be achieved. An optimal point along the curve is selected where it is agreed that this is the

minimum proficiency level necessary for job entry level skills, knowledge and attitude (see Figure 1). This is usually the point where the person can be productive in the job assignment without endangering him or herself or the organization. Further refinements, expansion and enhancements of proficiency are accomplished at the work site, on the job or through individual activities outside formal class work. This includes work experience, on-the-job training, mentoring and other formal and informal arrangements for instruction.

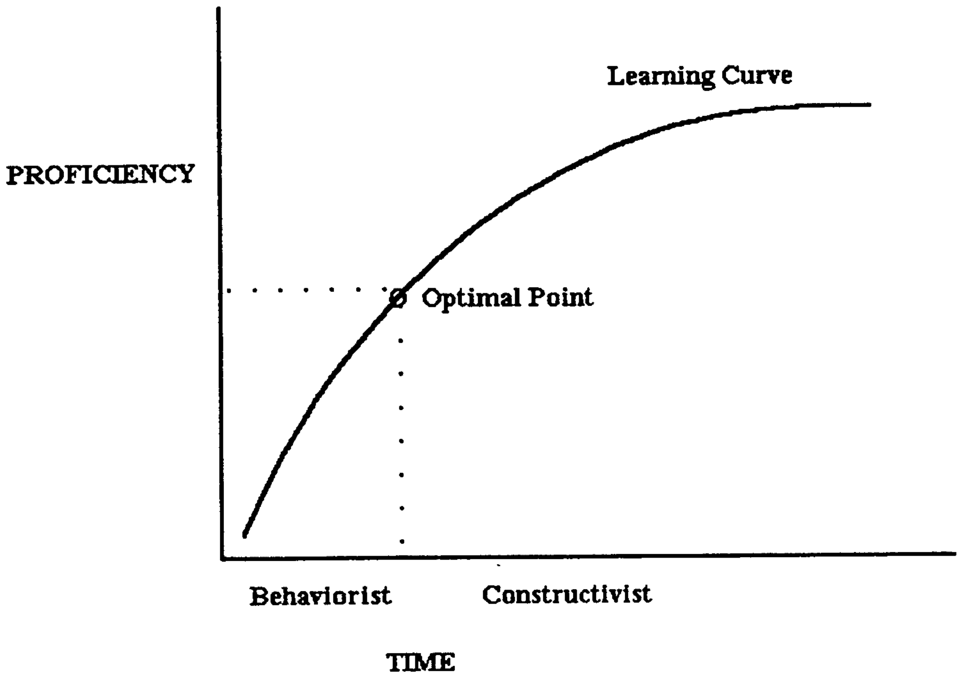


Figure 1. An Integrated Behaviorist/Constructivist Model

Distance learning technology offers a new tool for just-in-time, in place training. Advanced classes and modules can be offered via a PC based learning station in a classroom at the work site or on an office desk where workers can select instructional modules or courses that meet their particular learning needs. The model offers a structured approach for generic or basic skills (behaviorist approach) with the opportunity for acquiring additional knowledge and skills as needed by the individual (constructivism). The system could serve multiple functions in some organizations providing technical consulting service for operation and maintenance problems, technical library resources for research needs as well as training and education delivery. Indeed these functions overlap to great degree and are characteristic of the learning organization.

In this model, constructivism may be a powerful model for empowering and motivating the learner to continue his/her progress along the learning curve. In other words, the basic skill and their level of performance are highly structured behaviors where some sort of consensual standard has been established. After this formal, structured learning process is completed with entry level competencies certified, the learner is supported in continuing the learning process. It is in this stage where learning curve may be shaped by the learner to reflect individual needs and motivations. For a more detailed discussion of constructivism in instructional design see Jonassen (1990, 1991a, 1991b) and Merrill (1990, 1991).

Issues

There are a number of factors that need to be considered in the design and development of distance learning classes. The following list is by no means exhaustive. Each situation is unique and may include a varying number of the factors listed as well as others not identified. This provides a starting point for the analysis required.

Scheduled v. On-demand Instruction: In traditional standup training programs, classes are scheduled and students select or are selected for attendance according to a catalog produced by the organization providing the training. Administratively, it is necessary to accommodate facilities and resources to organizational training needs in some efficient and ordered manner. As a result, the instructional design must include all the knowledge and skills required for the trainee to be certified for particular job tasks to be performed over the next several months or years. These are usually entry level performance requirements and graduates are expected to hone their skills through on-the-job training, apprenticeship or mentoring. Much of the knowledge presented in the course content may not be encountered routinely by the trainee because it is infrequently required or because it is performed only by advanced practitioners. It is included in the curriculum because it is impractical to bring students back to the classroom for instruction just as it is needed. Distance learning systems, particularly desktop systems, offer an opportunity to provide just-in-time training and refresher instruction on demand. The Navy has found, for example, that even regular distance learning courses can be used by commands to provide refresher training for specific tasks. Diesel mechanics on duty with the fleet in one instance will be able to register for only the portion of a course they need for refreshing their knowledge on Diesel Operation and Maintenance that is being telecast from the Service School Command, Great Lakes Naval Training Center (NTC) (Larson, 1996).

In another case, crewmen on board a Navy Frigate, in San Diego were receiving distance learning instruction on gas turbine engine maintenance via a desktop two-way video, audio and data link from Service School Command, Great Lakes NTC, via Damneck, Virginia when an instructor noticed from their body language cues that they were unfamiliar with an oscilloscope test instrument needed for a maintenance procedure. A departure from the course curriculum was made for a class on the test instrument which solved a specific maintenance problem with the engines of their ship (USS Rentz Ship to shore, 1996). This example also illustrates the issue of two-way video v. one-way video with two-way audio and of real-time synchronous v. asynchronous interactions.

Two-way v. One-way Video: The traditional classroom offers face-to-face interaction within a full contextual frame. Both instructor and students see and hear the content along with gestures, eye-contact and the carriage of both students and instructor. The subtlety of cues is often so automated we are unaware of them or their effects on a conscious level, yet the dialog is adjusted to accommodate them. It is no accident that most comedy programs on television are produced before a live audience. The actors and crew need the feedback for their timing and inflection. The face-to-face arrangement provides a naturalness to the interaction that is not achieved through verbal range alone. Also, it is the familiar method of receiving instruction. Although there is little evidence in the literature that there is sufficient increase in learning efficacy to support the added costs of two-way video, most of the studies are methodologically flawed. No studies are found where the independent variable of two-way v. one-way video has been satisfactorily isolated from contaminant variables (Wetzel, et al, 1993). Certainly for some types of subject content involving skills training, it is intuitively evident that the ability to "see" student performance would be a powerful addition to the instructor's ability to allow practice and critique performance.

Evaluation and Assessment--Knowledge v. Skills: One of the reasons distance learning has been so widely and successfully used by colleges is that education courses are designed primarily by content analysis while military and industry training is more often based on task analysis. Whether content or task analyses are used does not change very much what is taught but it has great impact on how the

instruction is presented. Content analyses results in a cognitive domain content with performance measured, by and large, by recall of facts, concepts and constructs. Task analysis results in instruction that is activity oriented with achievements measured by task performance usually demonstrating some level of a skills mastery. Paper and pencil tests provide weak performance validity for skills measurement, but are quite adequate for knowledge acquisition. The distance learning designer/developer must select criterion reference test items that can be accommodated by the distance learning system.

Performance measurement, or testing, is perhaps one of the greatest weaknesses in education. In a study of test construction by college faculty, Main (1990) found that on average 85% of course grades in a college communication curriculum were knowledge based and measured by objective or short answer exams. Even in courses where students were required to use equipment and generate a product (audio production, for example), 60% of the course grade was based on objective tests rather than the authentic assessment measures. And, while objective exams can be constructed that test critical thinking and cognitive skills, the vast majority of items are simply recall of course content. Professors unanimously indicated they modified their exams to reflect only material covered during class periods.

Adult vocational training is much more dependent on task performance for learning assessment. When adapted to distance learning scenarios, accommodations must be made to insure the validity of skill performance measurement is maintained. If two-way video is available on the system, a "show me" exercise can be designed. When this is impossible because of time or system limitations, qualified evaluators at the distant site location can be employed to administer or monitor performance tests. Simulations (both computer-based and paper-pencil) can be created with problems and scenarios that will assess both procedural and problem solving skills. If qualified examiners cannot be assured at the learning site, the course designers must create other methods for recording student performance for evaluation. Students may be required to perform a writing task that demonstrates a skill and knowledge ability, proctors may video tape a physical performance (modern dance, for example), and portfolio reviews may be used where a collection of student products are assessed for competency level. These can be transmitted by digital file (e.g. e-mail, Internet, or the distance learning systems data links), by fax or by snail mail.

Test Administration

Two factors that contribute to student satisfaction are the opportunity to apply knowledge learned and the prompt return of assignments and tests (St. Pierre and Olson, 1991). In a distance learning environment, extra attention must be paid to the evaluation of learning outcomes. Because students are usually isolated and have limited opportunities for comparing their progress with others in the class, the frequency of evaluations may need to be increased and feedback on performance provided promptly. Performance ranges and test means should be available for students. Even when the training is competency based with no grades assigned, students want to know their relative performance with others in the course.

In a study by Cole, et al (1986), students in distance learning classes expect:

- -fair and objective grading;
- -to have their work treated with respect;
- -an explanation and justification for the grade awarded;
- -a clear indication of how to improve their performance;
- -encouragement and reassurance about their ability and progress;
- -constructive criticism and advice;
- -an opportunity to respond; and
- -a timely response (before the next assignment is due).

Test security presents a challenge in some situations. Proctors may be designated to administer the exam to students or on-line exams can be administered with students responding through their computer link in much the same fashion they would complete a quiz in the traditional classroom. Open book exams with open-ended and essay responses provide a reliable methods of evaluation.

Learning Activities--Individual v. Collaborative

Most educational learning activities are individual. However, in seminars, workshops and particularly in skills training courses, collaborative efforts are often desired. The ability of the distance learning system must be carefully examined to determine the capabilities that exist for group interactions. It may be possible to provide telecommunication links between students from multi-point sites which would provide the same interactions between students that exists between student and instructor. This capability requires advanced switching hardware and software that is just beginning to become commercially available.

The Internet provides the most universally available capability for students to work together. Text, visuals and even audio can be exchanged either asynchronously in e-mail accounts, bulletin boards, and files, or in real-time exchanges in chat rooms. established for the class or for student groups. If distance learning sites are in actuality satellite classrooms, students can do collaborative projects within their proximate group and report their results to the full class. The activities can be conducted during scheduled class periods or the work conducted during non-class time.

A course at California State University, Chico is being taught over the Internet exclusively with all assignments posted to the bulletin board and discussions held in chat rooms. Content is accessed from a website and research is conducted using data bases accessed from the Internet using a commercial search engine. Office hours are held on a scheduled basis by the instructor and by appointment. Both public and private interactions are possible. Student projects, individual assignments, even exams are distributed and turned in via the Internet. Administrative functions such as scheduling, registration adds and drops have been attempted but are not entirely successful to date. Payment of fees and access to other student services are not provided over the telecommunication network. Because of the widely divergent skills of students entering the class, a diagnostic pretest is administered to each student at the beginning of the semester. Students are ranked by score and paired as teams throughout the course. The top scoring student is paired with the lowest scoring student and the process repeated until all students are assigned into two person teams. The objective is to provide inexperienced students with a tutor teammate who can provide one-to-one assistance in learning the technology. Grades are team-based so that experienced students have an incentive to make their colleague proficient.

The University of Phoenix has eight years of experience in computer-mediated on-line education. They have found that when faculty take the time to orient their distance education students on self-direction and peer reliance, they can effectively diminish the teaching load as students themselves take more responsibility for meeting their learning goals (Hedegaard, 1996).

Contingency Planning

Although distance learning technologies are becoming increasingly reliable, they are subject to partial or total interruption. Strategies must be planned in advance and preparations made for contingency delivery of instruction. For partial system failure where one or two or even more students lose their communication link, a recording of the presentation screen and audio of the class can be made as a routine procedure and copies sent later. This is a relatively inexpensive and simple solution for keeping students on track, but care should be exercised that students do not abuse it to miss scheduled classes for personal convenience.

Most subject matter is hierarchical in nature, but their are components to every curriculum that do not require teaching in sequence. The instructional designer (with the instructor) should go through the lessons carefully designating learning activities that are dependent, those that are supportive and those that are independent of prerequisites. For example, teaching multiplication is dependent upon first knowing addition. On the other hand, computing the area of a rectangle is not a prerequisite for computing the area within a triangle. But, knowing how to do one helps in learning the other, i.e., some

of the knowledge is transferable. It doesn't matter which concept is taught first, but having learned one skill, it will take less time to teach the second. Learning to compute the area of a circle is, however, independent of computation for areas of triangles or rectangles. It doesn't matter which is taught first. By identifying these "independent" learning objectives, the instructional designer can prepare learning modules of these lessons in advance and have them packaged with other materials relevant to the students and sent to learners when they register along with instructions of how to proceed if a class is canceled.

Content may be prepositioned at the distant learning site(s) in a variety of forms and media. **Print** may be in hard copy, a floppy disc, CD-ROM or a data file on the hard drive or available on a server. **Audio** may be stored on cassette CD or as data file on a server. The same holds true for **video** segments. They can be prepositioned as video cassettes, CD-ROM's or, in compressed form, as data files on a server. **Computer Based Instruction (CBI)** can be available at the distant learning site stored on floppy disc, CD-ROM, or as a data file on the hard drive or on a server. Programs can be interactive multimedia modules, simulations or reference files. The instructional activities can be as varied as reading assignments, research and writing projects or interactive learning modules complete with competency based tests. Learning activities conducted outside the regular class period are often where higher order cognitive skills are required--analysis, synthesis and evaluation--as well as remedial and/or drill and practice exercises that reinforce knowledge acquisition.

Perhaps most important of all, using a mixture of media, allows for differences in student learning styles. Some learners prefer the reflective thinking associated with print. Others may be motivated by the competitive nature of an interactive game-based module or the concreteness and realism of motion video. The more media alternatives provided, the more effective the distance learning environment is likely to be for a wider range of students (Moore and Kearsley, 1996).

There are, of course, no assurances distant learners will actually use the contingency lesson materials anymore than they will accomplish homework or other out-of-class learning activities. To insure learners understand their responsibility for interrupted classes, the course syllabus or student learning guide should include instructions about the contingency lesson materials and how they are to be used. Test items, quizzes and graded assignments can be used to enforce student compliance if this is necessary.

Media Integration and Presentation Control

In most distance learning environments a combination of media forms is used. The benefits from a mixed media environment are many. No single medium can effectively meet all the learning objectives across a full course or program, the differing learning styles of individual students, or the capabilities of the delivery technology. Multiple media provide interest and flexibility. There are a number of very helpful models available to assist the course developers in media selection . Bill Walsh (1996) has developed an excellent summary of them in a practical guide for distance learning designers (App. B, in press).

Of more interest in this study is how the media will be presented. More specifically, will media presentation be centrally launched from the instructor's site or will individual learners exercise presentation control. The traditional method, of course, is instructor control. The instructor prepares the class with the use of the media scheduled in the lesson plan. At the appropriate time the media are used and the instructor continues with the remaining learning activities. While this procedure is familiar and comfortable for the instructor, there are a number of reasons for changing to student control of media presentations. The most important is the level of complexity that is introduced for network distribution of media by the instructor. Some of the issues are: 1) control over learner workstations by a central server, 2) inability of one or more sites to receive the signal, 3) data storage requirements for multimedia, and 4) network data transmission requirements for full motion video of sufficient fidelity for full screen instruction (Main, et al, 1996).

It may be preferable to position media at the distant learner's site where it will be used upon the appropriate cue from the instructor. Some of the benefits are:

It facilitates contingency planning when technology failures occur as previously discussed.

It distributes knowledge throughout the system allowing greater independence for access by individual learners.

It allows for differences in learner experience, abilities and/or motivation in time on task. Some learners may need to review the materials because they lack background or interest to master the initial presentation. It is interesting that we would not think of requiring students to read text in unison because we know the reading speeds vary so greatly. Just because the presentation is fixed-pace does not necessarily mean it is processed at equal speed by all learners.

It permits a more heterogeneous mix of technology. This could be extremely important for public education or training consortia course offerings. By having the media in a variety of formats, students can request the format compatible with technology at their location.

It makes the learner a more engaged participant in the learning process by transferring responsibility from the instructor for accessing the learning content.

A wider range of media-based learning activities may be possible. For example an interactive CD-ROM-based simulation may not be suitable for centralized distribution. It may be restricted to one or two participants at a time. Just as reading assignments are not appropriate for in-class activities, some interactive multimedia programs may also be best used off-line.

There are some potential disadvantages inherent in transferring control for media presentation to the learners (not the least of which is the learner's ability from both a technical and a psychological standpoint). It requires a certain level of confidence and self-discipline to discharge this responsibility as well as the technical skill to operate the software and hardware involved. Technical support may not be readily available if assistance is needed. Frustration thresholds can be very low when technology is involved. On the other hand, these skills and attributes may themselves be learning objectives which can be designed into the lessons as learning activities so that learners can be monitored and assistance provided as necessary.

The greatest advantage of local control (individualization of instruction) may, in some situations, be the greatest drawback. If standardization of the instructional *process* is important or if uniform progress is critical, then a lock step central control is preferable.

Summary

The overarching goal of instructional design for distance learning is that the technology be used to satisfy human needs and that human needs are not distorted to serve the technology instead. The technology should be used to enhance human to human contacts. If the instruction is intelligently designed with this in mind, the technology should tend to become transparent, not dominate the presentation (Bradley and Peacock, 1996).

Moore and Kearsley (1996) conducted an extensive review of distance learning literature and concluded distance learning requires:

-A greater emphasis on instructional design
-More instructor training than traditional classroom presentations.
-More money for instructional materials development..

As communication technology capabilities improve and education and training requirements expand, the need for alternative delivery methods is increasing.
The requirement to keep pace with the advanced technology and growth of knowledge in virtually every field, the changing nature of jobs and the increasing migration of workers between jobs and careers are some of the pressures for developing new ways to deliver instruction on demand. As the use of computer-based training and distance learning technology increases, there is need for new models or modification of old models for designing instruction. The purpose of this paper has been to surface some of the factors that should be considered in the process of designing instruction for distance learning.

REFERENCES

Braden, Roberts (1996), The Case for Linear Instructional Design and Development: A Commentary of Models, Challenges, and Myths, *Educational Technology,* March-April, 5-23.

Bradley, John and Kent A. Peacock (1996), Connecting to the Global Classroom: Distance Education in a University Setting, *ED Journal,* Vol. 10, No. 2.

Capell, Peter (1995), *Report on Distance Learning Technologies,* Software Engineering Institute, Carnegie Mellon Univ., Pittsburgh, PA.

Cole, S., M. Coats and H. Lentell (1986), Towards Good Teaching by Correspondence, *Open Learning,* Vol. 1, No. 1.

Fulford, Catherine P. and Shuquiang Zhang (1993). Perceptions of Interaction: The Critical Predictor in Distance Education. *The American Journal of Distance Education,* Vol. 7, No. 3.

Dillon, Alicia, (1996), unpublished master's thesis, California State University, Chico, Chico, CA.

Guiding Principles for Distance Learning in a Learning Society (1996), American Council on Education monograph, May, Washington, D.C.

Harasim, L. M. (1990), *On-line Education: Perspectives on a New Envizonment,* Praeger, New York, NY.

Hedegaard, Terri (1996), Computer-Mediated On-line Education: Lessons Learned by the University of Phoenix, *ED Journal,* Vol. 10, No. 2.

Jones, B. F., G. Valdez, J. Nowalkowski, and C. Rasmussen (1995), *Plugging in: Choosing and Using Technology,* Council for Educational Development and Research, North Central Regional Educational Laboratory, Washington, D.C.

Jonnason, D. H. (1990), Thinking Technology: Toward a Constructivist View of Instructional Design, *Educational Technology,* Vol. 30 No. 9, 32-34.

Jonnason, D. H. (1991a), Evaluating Contructivist Learning, *Educational Technology,* Vol. 37, No. 9, 28-33.

Jonassen, D. H. (1991b), Objectivism versus Constructivism: Do We Need a New Philosophical Paradigm?, *Educational Technology Research and Development,* Vol. 39, No. 3, 5-14.

Larson, Orv (1996), Personal interview conducted at the Naval Personnel Research and Development Center (NPRDC), San Diego, CA.

Main, Robert G., Terry Curtis, and Mick Presnell (1996), *Report of Findings Interactive Multimedia Distance Learning Experiment*, Darlene Hinds, ed., California State University. Chico, Chico, Ca.

Main, Robert G. and Eric Riise (1995), *A Study of Interaction in Distance Learning,* U.S. Air Force Technical Report, Armstrong Laboratory, Brooks Air Force Base, TX.

Main, Robert G. (1993). Integrating the Affective Domain in the Instructional Design Process, *Educational Technology,* Vol. 33, No. 12, 37-41.

Main, Robert G., Kate Robinson, and Joyce Scott (1992), *Integrating Motivation in the Instructional Design Model,* Air Force Office of Scientific Research Technical Report, Bolling Air Force Base, Washington, D.C.

Main, Robert G. (1990), *Instructional Design Practices Among Professors at a State University,* unpublished monograph, California State University, Chico, Chico, CA.

Main, Robert G. (1986), *Math Avoidance as a Factor in Career Choice Among Freshmen Communication Students,* unpublished monograph, California State University, Chico, Chico, CA.

Mason, R. and A. M. Kaye (1989), *Communication, Computers and Distance Education,* Pergamon Press, New York, NY.

Merrill, David., Z. Li, and M. K. Jones (1990a), Limitations of First Generation Instructional Design, *Instructional Technology,* Vol. 30, No. 1, 7-12.

Merrill, David M. (1991), Constructivism and Instructional Design, *Educational Technology,* Vol. 31, No. 5, 45-52.

Moore, Michael G. and Greg Kearsley (1996), *Distance Education: A Systems View,* Wadsworth Publishing Co., San Francisco, CA.

Nicelescu-Maier, Stefan (1995), *Grade Expectations as a factor in Student Evaluation of Faculty,* unpublished research paper, California State University, Chico, Chico, CA.

Salomon, G. and R. E. Clark (1977), Reexamining the Methodology of Research on Media and Technology in Education, *Review of Educational Research,* Vol. 47, No. 1, p 99-120.

Simpson, Henry H., Lauren Pugh and Steven W. Parchman (1991), *Empirical Comparisons of Alternative Video Teletraining Technologies,* (NPRDC-TR-92-3), Navy Personnel Research and Development Center, San Diego, CA.

USS Rentz Ship to shore (1996), video program of a training course from Great Lakes Naval Training Center to a ship at sea, SEAADSA DSN 332, Naval Sea Systems Command, Great Lakes, IL.

Wagner, E., (1990), Instructional Design and Development: Contingency Management for Distance Education, in *Contemporary Issues in American Distance Education,* Michael Moore (Ed.), Pergamon Press, p. 298-311.

Walsh, William J., (in press), Instructional Design Handbook for Interactive Distance Learning, National Center for Manufacturing Sciences (NCMS), Ann Arbor, MI.

Wetzel, C. Douglas, Paul H. Radtke and Hervey W. Stern (1993), *Review of the Effectiveness of Video Media in Instruction,*(NPRDC-TR-93-4), Navy Personnel Research and Development Center, San Diego, CA.

TIME-TO-CONTACT JUDGMENTS IN THE PRESENCE OF STATIC AND DYNAMIC OBJECTS: A PRELIMINARY REPORT

Philip H. Marshall, Professor
Ronald D. Dunlap, Doctoral Candidate
Department of Psychology

Texas Tech University
Lubbock, TX 79409-2051

Final Report for:
Summer Faculty Research Program
and
Graduate Student Research Program
Armstrong Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, DC

and

Armstrong Laboratory

August 1996

TIME-TO-CONTACT JUDGMENTS IN THE PRESENCE OF STATIC AND DYNAMIC OBJECTS: A PRELIMINARY REPORT

Philip H. Marshall, Professor
Ronald D. Dunlap, Doctoral Candidate
Department of Psychology
Texas Tech University

Abstract

The accuracy of time-to-contact (*TTC*) judgments in computer-generated visual displays was investigated in conditions that included no, static, or dynamic (moving) non-target stimuli. The number of such stimuli, and their direction and relative speed of movement also were manipulated. Analyses indicated that our tasks yielded traditional *TTC* functions, with undersestimation increasing as actual *TTC* increased (2-, 4-, 8-sec). The direction of non-target stimuli movement influenced *TTC* judgments only when they traveled at the same speed and in the same direction as the target. This effect was most pronounced at the longest *TTC*. Neither the number of non-target stimuli, nor non-target movement in general, affected *TTC* estimates. We suggest that a non-target stimulus may play several roles (have several influences) depending on the task requirements and the display configuration. Ordinarily one would think of non-target stimuli as distractors, but we suggest that when a non-target stimulus moves in the same direction and at the same speed as a target, it can assume the role of a "surrogate target," providing visible cues with which to judge target *TTC*. Within the limits of the conditions of this study, we conclude that *TTC* estimates are very robust, and are not easily influenced by otherwise extraneous variables, including accidental and potentially adverse testing environments. Performance on a *TTC* task, however, also may be determined by the adaptive nature of general strategic cognitive processes. We propose further research to determine if, when, and how extraneous stimuli may influence *TTC* accuracy, and what other adaptive and non-automatic processes might be inloved.

TIME-TO-CONTACT JUDGMENTS IN THE PRESENCE OF STATIC AND DYNAMIC OBJECTS: A PRELIMINARY REPORT

Philip H. Marshall and Ronald D. Dunlap

Introduction

For some time there has been a research interest in the ability of human observers to make time-to-contact (*TTC*) judgments. In one common version of this task, an observer watches a target traveling horizontally (at constant velocity) along a path for several seconds before that target disappears. The participant is to predict (usually by pressing a button) when the target would reach a predetermined end point or finish line. Typically, performance is characterized by increasing underestimation of *TTC* (responding earlier than the target would have made contact) as actual *TTC* increases (Schiff & Detwiler, 1979; Caird & Hancock, 1991). Some esearchers have suggested this ability to be solely a function of information from the optic array (Lee, 1976; Tresilian, 1991), while others have suggested the involvement of various cognitive processes and mechanism such as memory, imagery, and internal clocks (see Tresilian, 1995).

The stimuli in most *TTC* tasks consist of simple, moving objects (e.g., a square) in uncluttered displays, with no other stimuli. There are attempts currently underway to assess some potential distractor effects (Jennifer Blume, March 6, 1996; Gregory Liddell, May 6, 1996), and one published study (Lyon & Wagg, 1995) reports limited non-target stimulus effects with a target moving in a circular path. Research incorporating potentially distracting or other stimuli in the visual field can make contributions in several ways. First, real world situations in which *TTC* judgments are made are very likely to contain distracting or other events, and this is so even if the "real world" task is only monitoring a computer display. Therefore, research incorporating non-target stimuli is somewhat more "ecologically valid" than that where only a target is present and moves. Such research could also contribute to the debate on the extent of involvement of cognitive processes in *TTC* decision tasks. Cognitive acts that require effort (as distinguished from those that have become automatic) require a share of our limited attentional resources. To the extent that *TTC* processing is effortful, sufficiently distracting events could reduce attentional resources and affect *TTC* performance. Alternatively, there are other perceptual phenomena that might affect *TTC* accuracy when other stimuli, especially moving stimuli, are present in the visual array, and an example would be the so-called motion repulsion effect described by Marshak and Sekuler (1979). They found that the

perceived direction of motion for a given dot can be affected by the motion of another dot in the visual array such that the perceived difference between their respective headings is exaggerated.

In the present study, the presence, number, and direction of moving non-target stimuli were manipulated to determine possible effects on the perception of either the target's speed or path that would affect the accuracy of *TTC* judgments. It is worth noting that the nature of the effects of non-target stimuli could be to move the *TTC* function closer to actual times, that is, compensate for the underestimation normally observed. So, it would be naive to assume that the effects of the presence of non-target stimuli should always be in the direction of decreased performance, and we recognize that stimuli may have various functional roles depending on the situations in which they are present.

Method

Design

The variety of trials (stimulus scenes) in this study included those on which no non-target stimuli were present, those on which non-target stimuli were present but did not move (static), and those on which non-target stimuli were present and did move (dynamic). When non-target stimuli were present they varied according to how many there were (4, 8 or 16), and, when they moved, they varied according to their velocity relative to the target (same, or +/- 50%), and their direction of movement (0-315 degrees in 45-deg, counter-clockwise increments)

Participants

A total of 44 Air Force recruits participated at the start of this study as part of their basic training requirements. All (but one) were right-handed, had normal or corrected to normal vision, and participated according to standard Air Force privacy and confidentiality procedures. Two different computer systems were used (see below) and five participants from each had their data deleted because the participants either did not understand or follow the instructions. These individuals were identified by having a very large number of repeated trials relative to the majority of participants. The final distribution included 8 males and 9 females having used a Dell® computer system, and 7 males and 8 females having used a Micron® computer system.

Materials and computers

The two-dimensional scenes were programmed to have a light gray background, black vertical start and finish lines positioned in the middle third of the screen, dark gray square targets, and somewhat lighter, square non-target stimuli (approximately 83-, 0-, 16-, and 39-% of "pure" white, respectively). So, the targets

were made darker to distinguish them from the stimuli. Brightness settings at all stations were equated by turning all monitors to the brightest level. This had an overall effect of reducing contrast, but still clearly retaining the distinction between target and non-target stimuli. In any condition, when the target and a non-target stimulus overlapped or intersected, the target appeared to be in front of the non-target stimulus. All paths "traveled" by the target had the same finish line, but the start lines varied (see Table 1), and all movements were from left to right.

Each scene came on and remained static (nothing moving) until the subject pressed the spacebar to initiate that trial. Initially, the target was entirely visible, its trailing edge at rest against the starting line. When the participant depressed the spacebar the visible target traveled for 2-sec before it disappeared.

The targets traveled at six different velocities (see Table 1 for specifications of distance, velocity and *TTC*), two different velocities and distances after disappearing for each of the three times to contact. The non-target stimuli traveled at one of three different velocities relative to the target depending on which condition the participant was in. One third of the participants saw the non-target stimuli moving at the same velocity as the target, one third saw them moving 50% faster than the target, and one third saw them moving 50% slower than the targets. On any given trial all the non-target stimuli moved in the same direction, and followed a path defined by degree of deviation from horizontal (in increments of 45-deg, counter-clockwise from the horizontal, left-to-right direction of 0-deg).

When they were present, there were either 4, 8 or 16 non-target stimuli, randomly positioned on the screen at the start of each trial. Initial non-target stimulus positions were determined by randomly choosing an x,y intersection from an imaginary 16x16 grid that filled nearly all of the viewable area on the computer monitor (inset about 2.54-cm on all sides), with the restriction than no x or y value was repeated. If and when amoving non-target stimulus "left" the screen, a new one immediately appeared and began to move at a location at the other end of an imaginary circular path around the screen. Figure 1 shows examples of scene presentations for 4, 8, and 16 stimuli, and also indicates examples of three of the eight different non-target stimulus movement directions.

The six-item *TTC* matrix (two different scene conditions for each of the three *TTC* durations, 2-, 4-, and 8-sec, as in Table 1) was crossed with the three levels of Number of non-target stimuli and the eight levels of Direction of movement, for a total of 144 trials. There were also two replications of the *TTC* matrix on which no

Table 1

Target and Stimulus Specifications

Overall distance (degrees of visual angle)	Velocity (degrees/second)	Time to Contact (seconds)
23.5	5.9	2.0
17.8	4.5	2.0
17.6	2.9	4.0
15.2	2.5	4.0
13.4	1.3	8.0
11.2	1.1	8.0

Length of "start" and "finish" vertical lines was 5.6 degrees of visual angle.

The target and distractor squares had sides of .7 degrees of visual angle.

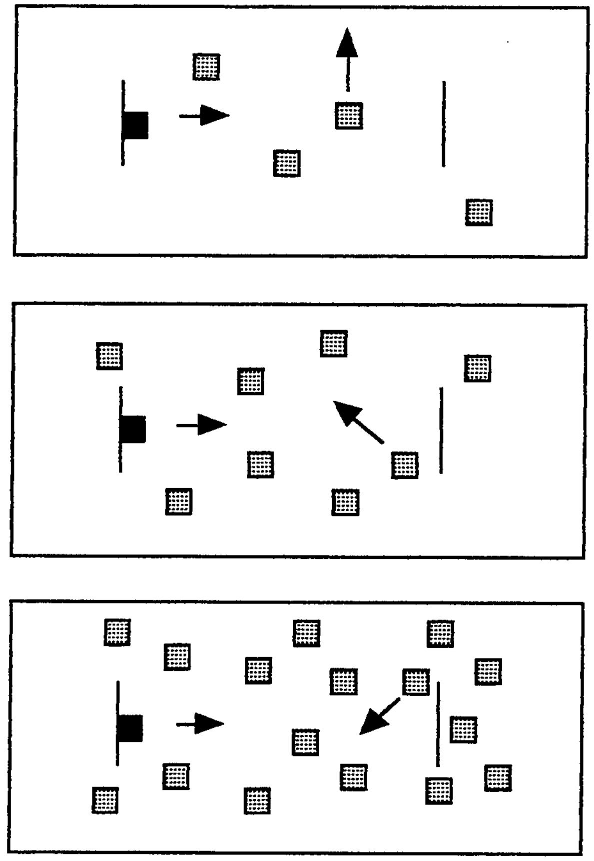


Figure 1. This figure shows samples of the three different numbers of non-target stimuli (4, 8, and 16), and three of the eight different directions of movement of the non-target stimuli.

non-target stimuli were present (12 trials), and six replications of the *TTC* matrix on which 2, 4, and 8 non-target stimuli were present (36 trials). Thus, there were a total of 192 unique trials.

In each session some participants used either a Dell® computer configured with a 90-MHz Pentium® processor with 16 megabyte of RAM, and a 17-in color monitor set to a black and white monochrome screen, or a Micron® computer using a 166-MHz Pentium® processor with 16 megabyte of RAM, and a 17-in color monitor set to a black and white monochrome screen The programs operated in EGA video, with a frame presentation rate of 14-msec per frame. We had no basis for predicting differences in performance based upon the systems, especially since frame rate was the same in both. In fact, a t-test on overall mean *TTC* estimates between the two systems resulted in an insignificant difference, 3.72-sec for the Dell® versus 3.81-sec for the Micron® [t (30) = -.31, $p > .75$], so the data from the two systems were pooled in the analyses presented below.

Procedure

The participants were run, on consecutive days, in two group sessions of 22 participants each. They were pseudo-randomly assigned to one of the computer stations, with the only restriction being that we attempted to evenly distribute men and women across computer systems and relative non-target stimulus speed conditions (normal, slower, and faster). The first part of the program described the use of the system, and demonstrated the stimulus conditions to be encountered during the study. There were also several practice trials with no non-target stimuli present, and which used a starting location longer than those used in the study, and a different (yet similar) velocity than any experienced in the study.

The presentation of criterion trials followed. To initiate each trial, the participant pressed the keyboard spacebar with left hand fingers to start the target moving, and pressed a mouse key using right hand fingers to make the *TTC* response. Upon the conclusion of the *TTC* response no feedback was given, and the scene for the next trial immediately appeared. The sequencing of the 192 trials was randomly determined for each participant. To compensate for inadvertent responses and possible inattention, a trial on which a *TTC* response occurred before the target had disappeared was aborted and was presented again at the end of the original series, as was any trial for which the *TTC* response was shorter than .5-sec. or longer than 12-sec. No trial was repeated more than once, and the average number of repeated trials was 13.35 (sd = 11.2), or just about 7%. Finally, participants proceeded at their own

pace with two one-minute rest breaks (remaining in their chairs and posturally oriented) after the 64th and 128th trials.

An important point needs to be introduced at this juncture. On day one of data collection there was an unplanned environmental occurrence, with the air conditioning in the testing center shutting down. Since the experimental sessions were conducted in mid-summer, the temperature and humidity in the testing center on that day became high enough to produce obvious general discomfort. Environmental data recorded in the testing center showed that the temperature had risen to 90°-F, with a humidity reading of 76%, sufficient to qualify for a "Category II" apparent heat index of approximately 110° F which can be associated with heat exhaustion in instances of prolonged physical activity (Steadman, 1979). Decrements in performance on visual processing tasks also have been found at this temperature (Hohnsbein, Peikarski, Kampmann & Noack, 1984). On day two of data collection the malfunction had been repaired, and readings were a much more comfortable 76°-F, with 72% humidity. In effect, we had an unplanned source of variance, a new factor - moderate heat-induced stress. This heat stress factor is introduced in the following analyses as the Day factor - high heat for day one, and normal conditions on day two.

Results

In each of the analyses that follow, mean *TTC* scores were computed over trials with actual *TTC* times of 2-, 4-, or 8-sec (respectively) in each condition, and those means were the data entered into the analyses of variance. Thus, for the no non-target stimuli and static non-target stimuli conditions, each *TTC* entry for each participant was based on four trials (observations), while in the dynamic non-target stimuli condition each *TTC* entry was based on two trials.

<u>Does the presence or mere movement of non-target stimuli affect *TTC* accuracy?</u> To answer this question *TTC* performance was assessed across the three task conditions with Gender and Day as between-subjects variables, and Task and *TTC* as within-subjects variables. That analysis yielded only an overall effect for *TTC*, $\underline{F}(2, 56) = 187.06$, $\underline{p} < .0001$, with mean estimated *TTC* increasing as actual *TTC* increased, 2.17-, 3.68-, and 5.58-sec for actual *TTC* times of 2-, 4-, and 8-sec, respectively. No other main effects or interactions were significant at the .05 level. Thus, the mere presence (static condition) or movement (dynamic condition) of non-target stimuli did not have a significant affect on overall *TTC* estimates relative to the simple condition where only the target was present.

<u>Does the number of non-target stimuli present have an effect on *TTC* accuracy?</u> To answer this question an analysis of variance was performed on data from the static and dynamic conditions. In the latter, performance was pooled over

the direction of movement manipulations. This analysis had Gender and Day as between-subjects variables and Task, Number of non-target stimuli (4, 8, or 16), and *TTC* as within-subjects variables. There was a significant effect for *TTC*, $\underline{F}$ (2, 56) = 173.13, $\underline{p}$ < .0001, with increasing mean *TTC* estimates of 2.18-, 3.69-, and 5.55-sec. There was also a significant interaction between Day and Number of non-target stimuli, $\underline{F}$ (2, 56) = 3.85, $\underline{p}$ < .05. Day 1 (Heat) participants gave slightly longer estimates of *TTC* than Day 2 (Normal) participants, especially for the eight non-target stimuli condition. Although we had no a priori hypothesis about the effects of heat, it might be that the high heat and moderate numbers of non-target stimuli combined to create an optimum arousal-optimal performance situation, but such an explanation is purely speculative, and, in any event, Day (testing temperature) did not interact with *TTC* duration. The number of non-target stimuli yielded no main effect, nor did that factor interact with *TTC*.

Do the number, relative speed and direction of movement of non-target stimuli have an effect on *TTC* performance? To answer this question an analysis of variance was conducted on data only from the dynamic condition. That analysis had Gender, Day, and Relative Speed of non-target stimuli as between-subjects factors, and Number of non-target stimuli, Direction of Movement, and *TTC* as within-subjects variables. That analysis yielded a significant main effect for *TTC*, $\underline{F}$ (2, 40) = 137.25, $\underline{p}$< .0001, with increasing mean *TTC* scores of 2.12-, 3.66-, and 5.58-sec. There also was a significant interaction between Direction of movement and *TTC*, F (14, 280) = 4.83, $\underline{p}$< .0001, and between Relative Speed of movement of non-target stimuli, Direction of movement, and *TTC*, $\underline{F}$ (28, 280) = 1.91, $\underline{p}$ < .01. This latter interaction, encompassing the effects of the former, is shown in Figure 2. Time to contact estimates increased as a function of actual *TTC*, but there was the usual greater degree of underestimation of *TTC* as actual *TTC* increased. Further, with non-target stimuli moving at the same speed as the target, participants were much more accurate in their *TTC* estimates at the longest *TTC* duration when the non-target stimuli traveled in the same direction as the target. In this instance, underestimation was virtually eliminated. No other effects were significant at the .05 level.

What is the nature of individual differences in *TTC* performance? To answer this question we constructed an individual differences variable representing overall *TTC* performance in each of the three conditions. We chose the slope of the linear regression line fitted to each participant's mean judged *TTC* for the three actual *TTC* values of 2-, 4-, and 8-sec in each condition. The distributions of slope values for each task condition are shown in Figure 4. A slope value of 1.0 indicates perfect accuracy in *TTC* ability, while values less than 1.0 indicate a tendency towards underestimation,

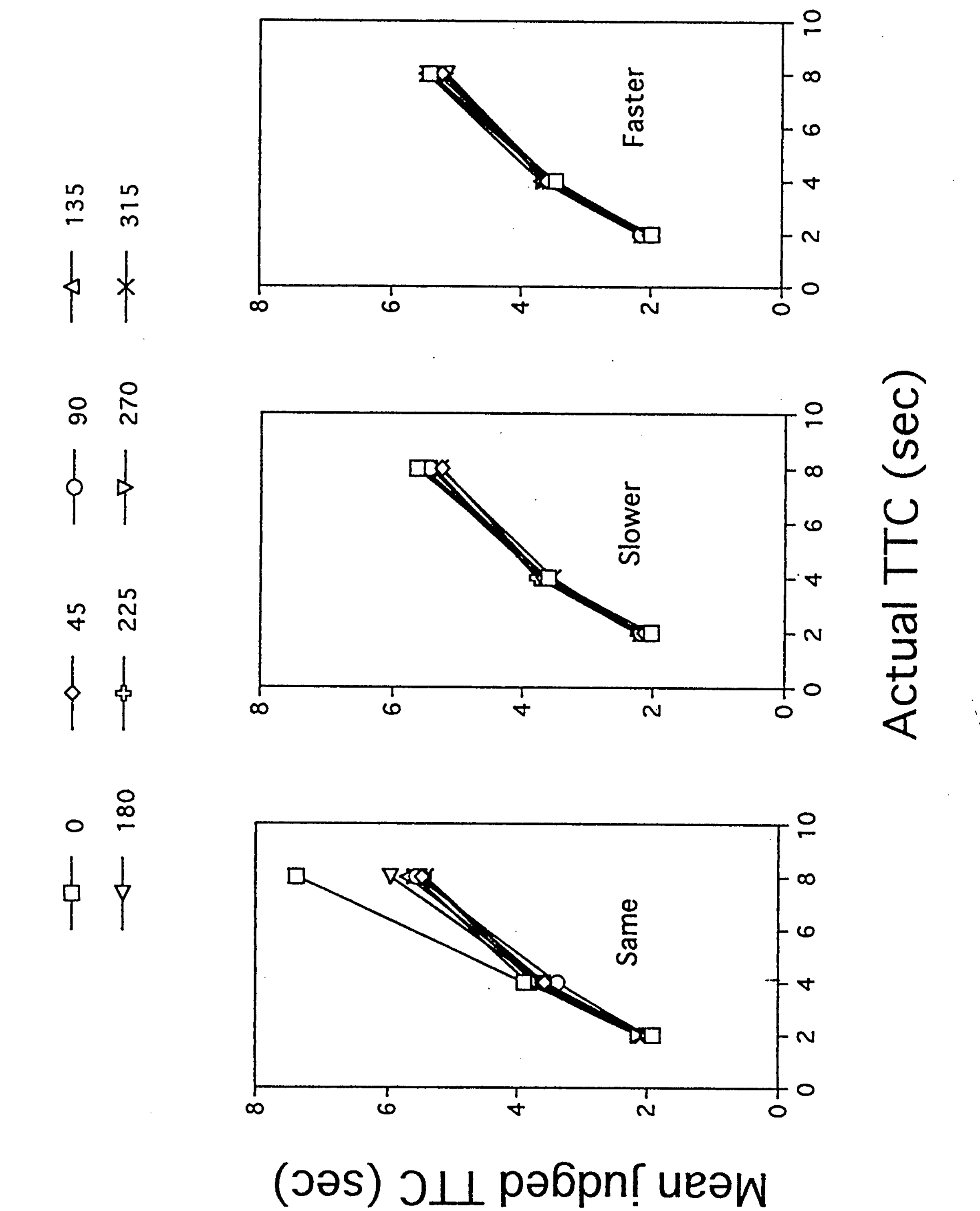


Figure 2. This figure shows the interaction between Speed and Direction of non-target stimuli and *TTC*.

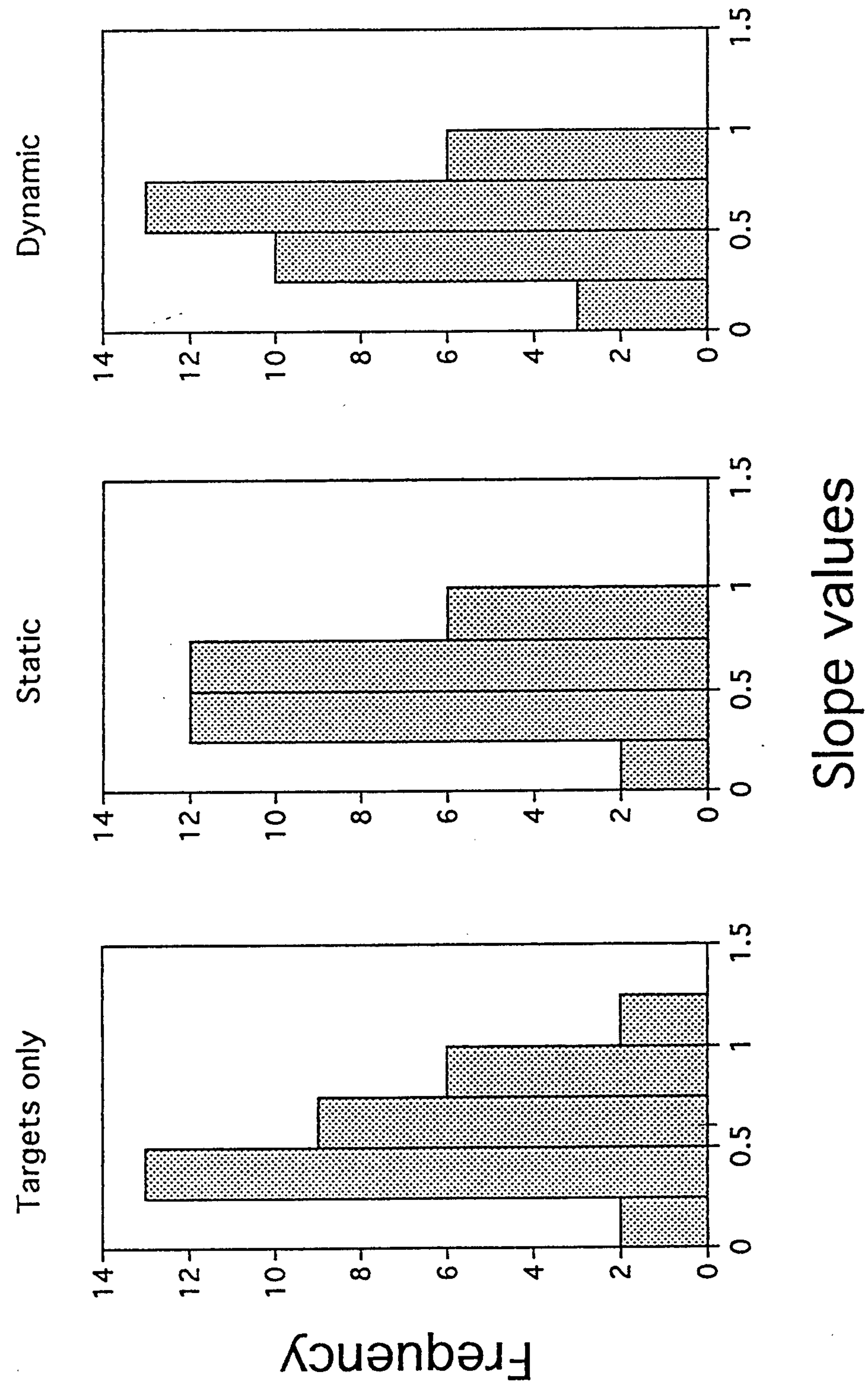


Figure 3. This figure shows the distribution of *TTC* slope functions under the three task conditions.

and values greater that 1.0 indicate a tendency toward overestimation of *TTC*. Nearly every slope value is less than 1.0, but one can observe a substantial range of values, and the possibility of an underlying normal distribution of *TTC* performance accuracy.

Discussion

We began this study with some expectation that the non-target stimulus manipulations would produce deleterious distraction effects on *TTC* performance, but we were not sure how those effects would be manifested in performance. It appears from our results, however, that *TTC* performance is rather difficult to disrupt. Non-target stimuli, even when they are numerous and moving across the target's path, do not seem to disrupt *TTC* judgments. We also had the opportunity to observe that not even a very hot and uncomfortable task environment produced a disruptive effect. In fact, the only substantial effect on *TTC* performance, other than the obvious effects of actual *TTC*, was the improvement in accuracy when the non-target stimuli moved at the same speed, and in the same direction as the target at the 8-sec *TTC*, but there is a plausible explanation for that facilitation effect.

A non-target stimulus traveling in the same direction and at the same speed as the target stimulus is essentially a running mate, and can become a <u>surrogate target</u> when the actual target disappears. One merely has to make a mental note of the degree of separation between the target and a correlated non-target stimulus, and use the movement and location of the surrogate non-target stimulus as a guide to when the target would reach the finish line. The longer the remaining travel time before the target would have contacted the finish line, the more time for the participant to make these mental compensations, and hence performance at the 8-sec *TTC* received most of the benefit of the surrogate process. Non-targets moving at different speeds or directions would serve the surrogate function less well, if at all, and that also is consistent with our findings. Unfortunately, we did not collect interview data following the task, so we have no direct confirmation of the surrogate process. This surrogate facilitation effect could be confirmed and further investigated in experimental situations in which, for example, the non-target stimuli themselves disappear sooner or later during the target's invisible period. The sooner they disappear, the less effective surrogates they would become.

Tentative acceptance of the surrogate explanation does confirm, somewhat, our initial speculation that the simple, target only, laboratory tasks could be overly simplified representations of conditions encountered in the real world. Apparently, in our much richer dynamic displays, our participants found a correlated (predictive) cue, the surrogate, to help them with the task. In fact, it may have

helped them so much that the usual increasing deviation from the actual *TTC* could be eliminated in some situations (e.g., the same velocity, 0-deg, 8-sec *TTC* condition). Further, real world situations are often replete with the presence of, and the opportunity to make use of, such "decision aids," so the surrogate effect should not be dismissed lightly. Rather, it should be recognized as being an instance of the use of general, strategic, and adaptive cognitive functions.

We considered that a motion repulsion effect might operate to influence the perception of the path of motion of the target, especially after it had disappeared from the screen. We found nothing to support motion repulsion effects in our data, and that is probably because the finish line was always visible to be a heading cue. In the absence of a visible finish line (one that disappears along with the target), or under conditions where the finish lines vary in direction, there may very well be a greater opportunity to observe *TTC* biases consistent with motion repulsion effects.

One of our original speculations was that non-target stimuli might consume some of the limited attentional resource available for the *TTC* task and decrease performance, but we have little to offer to confirm that notion. Attentional resources may not have been diverted by the non-target stimuli. Or, attentional resources may have been consumed by non-target stimulus conditions, but there may have been sufficient resources remaining to time share the *TTC* tasks. Or, *TTC* tasks may not require attentional resources to be performed. Hasher and Zacks (1979), among others, have suggested that, for humans, certain types of encoding operations require little or no attentional resources, and these have to do with the flow of information. While their concern was with memory operations encoding frequency, temporal order and spatial information, their general framework might extend to simple timing phenomena as well, since timing is essential in monitoring information flow. Indeed, there has even been speculation that the ability to time the release of projectiles, with the intent of hitting a stationary or moving object, might even be the basis for a hominid evolutionary drive (Calvin, 1983). Certainly, *TTC* estimation would be representative of such abilities, and it would serve evolutionary purposes well (not to mention individual survival!) if such abilities were not easily disrupted.

In the main, and under the limited conditions the present study, the results emphasize the robust nature of *TTC* decision operations. They are difficult to disrupt, and do not appear to be affected adversely by the presence of various numbers of non-target events, even when they move in a direction opposite to the target. Researching a related perceptual ability, Royden and Hildreth (1996) and others (Cutting, Vishton & Braren, 1995; Warren & Saunders, 1995) have made similar

observations with respect to research on heading judgments, concluding that "...moving objects do not significantly affect an observer's heading judgments in real situations..." (p. 851). Time-to-contact decisions may very well be based on similarly durable processes, but just how durable must be decided by further research.

References

Caird, J.K., & Hancock, P.A. (1994). The perception of arrival time for different oncoming vehicles at an intersection. Ecological Psychology, 6, 83-109.

Calvin, W.H. (1983). A stone's throw and its launch window: Timing precision and its imlications for language and hominid brain. Journal of Theoretical Biology, 3, 115-124.

Cutting, J.E., Vishton, P.M., & Braren, P.A. (1995). How we avoid collisions with stationary and moving objects. Psychological Review, 102, 627-651.

Hasher, L. & Zacks, R. (1979). Automatic and effortful processes in memory. Journal of Experimental Psychology: General, 108, 356-388

Lee, D.N. (1976). A theory of visual control of braking based on information about time-to-collision. Perception, 5, 437-459.

Marshak, W. & Sekuler, R. (1979). Mutual repulsion between moving visual targets. Science, 205, 1399-1401.

Royden, C.S. & Hildreth, E.C. (1996). Human heading judgments in the presence of moving objects. Perception & Psychophysics, 58, 836-856

Schiff, W. & Detwiler, M.L. (1979). Information used in judging impending collision. Perception, 8, 647-656.

Steadman, R.C. (1979). The assessment of sultriness. Journal of Applied Meteorology, 18, 861-884.

Tresilian, J.R. (1991). Empirical and theoretical issues in the perception of time to contact. Journal of Experimental Psychology: Human Perception and Performance, 17, 865-876.

Tresilian, J.R. (1995). Perceptual and cognitive processes in time-to-contact estimation. Analysis of prediction motion and relative judgment tasks. Perception and Psychophysics, 57, 231-245.

Warren, W.H. & Saunders, J.A. (1995). Perceiving heading in the presence of moving objects. Perception, 24, 315-331.

Acknowledgements

We thank the Air Force Office of Scientific Research for supporting this research, Dr. William Tirre of Armstrong Laboratory, Brooks Air Force Base, San Antonio, Texas for encouraging and facilitating all aspects of this project, and Ms. Karen Raouf, whose programming expertise was essential. We especially thank our families for cheerfully coping with our long absences during the conduct of this study. Comments may be sent by email to *p.h.marshall@ttu.edu*.

Sandra McAlister's report was not available at the time of publication.

ENVIRONMENTAL COST ANALYSIS: CALCULATING RETURN ON INVESTMENT FOR EMERGING TECHNOLOGIES

Bruce V. Mutter
Associate Professor
Division of Engineering Technology

Bluefield State College
219 Rock Street
Bluefield, West Virginia

Final Report for
Summer Faculty Research Program
Armstrong Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, DC

and

Armstrong Laboratory
Tyndall Air Force Base, FL

August 1996

ENVIRONMENTAL COST ANALYSIS: CALCULATING RETURN ON INVESTMENT FOR EMERGING TECHNOLOGIES

Bruce V. Mutter
Associate Professor
Division of Engineering Technology
Bluefield State College

Abstract

This research examines the process of calculating the Return on Investment (ROI) for emerging technologies. The report illustrates the relationship between means and costs associated with implementing appropriate technologies to solve a compliance, remediation or source reduction problem. Major cost factors were identified by comparing emerging technologies to a baseline capable of achieving equivalent end results. The range of the costs captured for each alternative was developed by a decision criteria model and included: direct, indirect, liability, and intangible costs. The tabulation of total costs was input to conventional Net Present Worth (NPW), Internal Rate of Return (IRR), and Benefit Cost Ratio (BCR) equations, which were presented to solve for the time value of the total cost estimates. Finally, the Return on Investment (ROI) was calculated for an emerging technology based on results of the life-cycle cost estimate.

ENVIRONMENTAL COST ANALYSIS: CALCULATING RETURN ON INVESTMENT FOR EMERGING TECHNOLOGIES

Bruce V. Mutter

Introduction

The inherent limitations of conventional cost analysis become apparent when assessing the financial performance of an investment in emerging environmental technologies. These limitations appear to be widely recognized. We now know that some short-sighted *non*decisions were made, in business past, because we were ignorant of the true costs. If we had known of our future obligations to comply, the cost of our liabilities, then we would have probably taken those costs into account and made more informed decisions. We now have the opportunity to learn from this past to make better investment decisions in the future. Organizations will to return, again and again, to reinforce the fact that all business decisions will be forever linked to the cost of one option compared to the cost of another. The difference, from now on, should be that the costs are more accurately allocated to a process or activity, and are reflective of the true cost of doing business. How?

The practice of merging and comparing the success of environmental processes with asset, resource, income, cost and managerial data is a complex process that defies convention. Accounting for the full cost of project alternatives is not an "easy sell" in any organization. When environmental costs are identified and quantified, the direct (capital, operating, and regulatory), indirect (training, audits, fines, etc.), and intangible (contingent, liability, good will, etc.) costs provide insight into the real cost of the effort (Kirschner, 1994: p.25). Certain limitations of conventional cost analysis surface immediately. Uncertainty in quantifying environmental costs is the primary cause for these limitations. We should focus on minimizing the uncertainty associated with calculating the return on investment.

Environmental research will continue to yield innovative solutions, but the questions remain: What is the nature of the true environmental costs? How large could the liability costs be? When will the costs occur? If investments in innovative control, compliance, and remediation methods are in the interest of the organization, then what accounts for the reluctant approach to invest the capital in new solutions? How can management afford to invest in new

programs in an atmosphere of limited resources, where these projects compete with more pressing mission tasks (White et al.,1991)?

There are a few readily apparent explanations for the contradiction drawn between environmental policies and the final financial analysis used to make the decision to carry the policies forward. First, available alternatives might be precluded from the decision making process, from the outset, due to organizational structure or the general attitude toward progressive environmental projects. In this case, there will be no cost analysis. Second, if the decision to undertake a new environmental project has been made, then the not-so-obvious economic or financial barriers tied to methods of costing or the budgeting process block the inclusion of the total cost and/or the proper life-cycle for the project in question (White et al.,1992-93: p. 35). Third, there is a general lack of credibility or stigma attached to liability costs and less tangible qualifiers, which can lead to their exclusion. These costs are often referred to as "externalities" for this reason. Evidence has shown that externalities can rather quickly become internal costs under certain circumstances. Can we afford to completely exclude these future costs from the evaluation of our options?

All organizations, whether large entities or small businesses, face the dilemma of calculating how to appropriate scarce resources to competing projects. However funded, most environmental control, compliance, or remediation projects are subject to some sort of profitability analysis. This process is used to assess the desirability of one project over another, or the break-even point established by the organization for undertaking a project. Organizations need to examine innovative environmental cost analysis techniques that measure up to the task of calculating Return on Investment (ROI) for current and emerging environmental technologies. New ROI strategy will be founded on capturing the total costs throughout the life-cycles of the emerging technologies versus continued use of current methods. For some environmental projects, there will be no return on investment. In these cases, we can devise a system to consistently select the alternative that minimizes the cost of what we must do anyway. Within the context of the corporate budgeting structure, we should determine if, and to what degree, conventional methods of investment analysis act to distort the cost of innovative methods in favor of more conventional measures (White et al., 1991: p.9).

Since we are, in essence, determining how to best allocate funds, the organization's budget becomes one of the constraints. Budgeting is a strategic process of analyzing alternative investments and deciding which one is best for the organization. The nature of budgeting for

capital expenditures often requires a five, ten, or more years' outlook. This is important because it will indicate how far into the future the organization is willing to analyze costs or accept life-cycles of competing alternatives. Large corporate entities should primarily focus on long-term operations for calculating return on investment. Therefore, they require a formal budgeting process involving input form many departments within the organization (Todd, 1993: p. 3-9). This can cause the collection and dissemination of detailed cost data to be conflicting and laborious. However, the sophistication of the financial analysis should match the size and long range expectations of the organization and assembling this cost data is key to the success of the analysis. Small private firms are better able to focus on short term profit and more often than not make decisions based on simple payback calculations, but institutions with centuries old financial histories and long range plans that the future itself depends upon, can ill afford to use of such capricious decision-making techniques to manage their finances (White et al., 1991: p.10).

The challenge is to develop a framework that can be relied upon to consistently help make the best selection among alternatives, and help answer the question: where is the best investment of scarce capital resources for environmental projects? The technology alternative that offers the largest return or minimizes the financial burden relative to cost should be chosen. This involves quantifying; placing some dollar value, whenever possible, to risks and uncertainties that traditional cost analyses have not yet articulated. Rapidly changing regulations, and the court decisions that define them, continually alter costs. Determining risk associated with treatment, storage, and disposal facilities (TSDF) for hazardous materials further complicates the cost analysis. However, we should continually adapt environmental cost analysis in response to these complexities, so that we make the make the most informed decision possible and hopefully select the lowest cost alternative that will best serve our future needs.

Problem Statement

The development of new environmental technology requires investment decisions to be made in an atmosphere of limited resources. The competing project alternative with the greatest return on investment or lowest opportunity costs should be selected. The objective is to develop guidance for evaluating emerging environmental technologies in comparison to a baseline that leads to the selection of the better investment. The comparison must take into account the total costs, while making sure the competing technologies can accomplish the same end result. A five step approach will be necessary to calculate the return on investment for emerging technologies, which consists of the following required elements: specify the environmental cost problem, develop a decision criteria model that allows comparison of the alternatives within the constraints of the model, capture the total costs of alternatives, apply the life-cycle to the cost estimates, and finally, calculate the return on investment. Case studies will be presented to illustrate the capabilities of this procedure in actual practice. These examples will demonstrate the return on investment for several emerging environmental technologies. The sample analysis will be capable of spelling out the costs and benefits of varying types of emerging technologies, at different stages of development, so decisions are more universally well informed.

Methodology

Specifying the Cost Problem

All environmental costing problems are of a common nature. These problems have two essential characteristics. First, there is a deviation between what managers think it will cost and the *actual* cost to solve an environmental problem. The second characteristic is that the deviation is important enough that the responsible decision maker thinks it should be corrected. The second part of the cost deviation is what makes it a problem. Why are excessive costs created? Deviation from the cost as planned could be brought about by an unanticipated regulatory change. Some of the other excessive cost generators may include: lack of cost information on new processes, myopic financial analysis, immature stage of technology development, lack of standards for performance measurement, and shortage of time available to analyze the costs and implement a solution. The implementation of new technology has a long term economic impact that is sufficiently important to be part of any analysis leading to a decision. There may well be a great many other aspects of the problem to consider before making a decision, but cost will dominate any decision process, and therefore, should be the focus of the problem specification (Newman and Johnson, 1995 p:4).

There will usually be several technologies, methods, or processes that could solve a compliance, remediation, or source reduction problem. The combination of attributes for each alternative can complicate side by side comparisons, turning the valuation process into an intricate matrix instead of a step by step approach. The objective is to identify the costing problem, and continue with problem formulation, including all relevant goals and objectives. This will eventually lead to the establishment of operating profile criteria. When specifying the cost problem, the analyst searches for those constraints that draw a boundary line around the relevant and important cost drivers. The cost problem can be properly specified by systematically answering the following questions:

What is the process in which the cost deviation was observed? Where is the process ? When did/will the excessive cost first appear? What is the amount of the cost deviation? What regulatory factors may have contributed to this cost difference (Macedo et al., 1978: p.42)? In a continuing process we need to recognize the environmental cost problem, state the specifics of

the problem, develop possible causes of cost deviations, then test the problem specifications, until satisfied that the true cost drivers have been identified. This will ensure that we have considered all relevant information, achievable goals, and feasible alternatives and filtered this through the organization, so that a reliable decision criteria model can be developed.

Developing a Decision Criteria Model

The technology will be selected from feasible alternatives. These alternatives need to be identified and then defined for subsequent analysis. When an emerging technology is compared to a baseline, then a set of decision parameters must be established for the scope of work involved. The context of comparison affects the end result, therefore, the analyst needs to determine how and where the new technology would be applied. Assessing the situation in such a way, as to realize the constraints (strengths and limitations) of the competing technologies. This is an important part of making sure the alternatives will meet common achievable objectives. We need to base the selection of competing alternatives on performance under the most probable set of conditions, then proceed with the cost analysis. The technology profile can be brought to light by such questions as: What is the nature of the contamination? Where are the projects located? What is the capacity of the assembly, device, or technology units that are to be installed? How many units or assemblies are needed based on their performance? How long would the technology need to be in place? What level of cost reporting detail is required to assess relative performance? If care is taken in understanding the big picture first, then properly detailed descriptions can be used to identify the components that make up the assembly of technologies (U. S. Army, EIWR,1995: p.38).

The purpose of establishing the decision criteria model is to synthesize the goals and objectives of the technologies, with the relevant cost and performance data available for both the baseline and the emerging. This will increase the awareness of cost drivers within the applications of the technologies. We should focus on the differences; only the differences in expected future outcomes among the alternatives are relevant to their comparison and ought to be considered in the analysis. For example, research and development cost data may not be relevant to the cause, if this cost data is not available for both the old and the new process. We want to measure the corresponding performance attributes of the alternatives. This objective can be realized by monitoring proper gauges of performance. We should measure enough parameters to indicate the range of capabilities for the technologies. If comprehensive standards are established, then the resulting indicators can be a useful tool to compare alternatives, recognize constraints of

available options, and in the context of one particular setting, determine whether the emerging technology is a functional alternative (Booth et al., 1991: p.13). An example of this approach could be applied to permeable barriers. This method is known to be an emerging technology. Profiling assemblies here would require comparing this passive (in situ) groundwater remediation technology to *other* treatment trains providing groundwater remediation solutions. It would be necessary to conceptualize the geometry and geography of the site, the nature and extent of the contamination, and the depth of the installation required to do this specific job. This would allow us to determine when an alternative is most effective.

The construction of the decision criteria model should ensure that the comparison will be technically consistent, and eliminate, or at least diminish the uncertainties of the cost problem. The intent is to compare practical applications of the competing alternatives and identify an approach that can be used to capture costs for further analysis. An alternative may drop out of consideration early because it offers little promise of meeting the performance requirements, or later on, because its relative cost is so great that it is not a reasonable cost alternative. We want to screen the approaches, and eliminate those options that do not fit the performance profile. An important tenet, at this stage of the process, is that we are working to develop scaled risk factors. The decision criteria model illustrates, whether or not, research and development of the emerging technology is mature enough to predict performance with a reliable degree of confidence (Showalter et al.,1995: p.22).

Using a common unit of measure will allow appropriate comparison of the innovative technology to the baseline, leading to the decision to fund additional research, development, testing, evaluation, and the eventual implementation. The emerging technology must be better than the accepted technique, both technically and economically, to compete for limited resources (Showalter et al.,1995: p.18). This comparison should be as objective as possible to avoid any conflicts of interest in the decision making process.

The crucial part of developing the decision criteria model is to make sure the techniques are truly comparable. A common reference point might be established by some type of performance standard; that is, competing projects should be capable of achieving the same level of pollution control or source reduction. There will be few cases where the emerging technology serves as a direct replacement for an existing one. The nature of innovative processes is that these new technologies usually do not fit with current practices on a point by point basis; rather,

they are a different means to the same end. Differences in capabilities will have to be articulated with the decision criteria model, which should incorporate real world analysis. For example, if the latest goal is to reduce the volume of exhaust requiring treatment of air contaminated with volatile organic compounds (VOC) passing through painting facilities by 90 percent, then a baseline technology that can reduce the volume by that amount would make an acceptable comparison. In addition, this particular case might call for comparison of two emerging technologies in tandem, in contrast to technologies that make up the conventional treatment train. The reason for pairing emerging technologies in the decision criteria model here is that recirculation in conjunction with a source leveling device, might be more cost effective than the independent application of each. It depends on how it fits into the rest of the treatment system. Therefore, we compare the accompanying emerging technologies to those baseline technologies that make up the current train of treatment, being cautious, all the while, that the competing assemblies will yield the same end result or that the cost of any compromise is included in the final analysis.

Capturing the Total Costs

While the relative performance capabilities of the alternatives are being evaluated, we must also examine the elements associated with total cost analysis within the context of implementing these technologies. For environmental technology projects, there are four general cost categories: (1) direct costs (2) indirect costs (3) liability costs and (4) intangible costs. Emerging environmental technologies warrant the capture of a wider range of cost components for consideration than conventional cost analysis projects have traditionally required.

As we will see, the decision to invest in new technology may not be possible based purely on *direct* costs. We should consider all relevant cost data, if we want to capture the *true* costs that go into the decision (EPA, 1989). If costs can't be expressed in dollars, then these costs should at least be made explicit with descriptive qualifiers to accompany the cost figures. In certain cases, the probable costs avoided should be considered, along with conventional capital costs associated with the investment. A further breakdown of these costs is discussed in turn.

(1) Direct Costs

Capital Expenses - Initial Acquisition

- *o* Site Work - land acquisition, surveys, file fees, clearing, relocation, drilling, & fencing
- *o* Construction - general conditions, substructure, superstructure, enclosure, & finishes
- *o* Equipment - conveying systems, HVAC, fire protection, security, process, & non-process, operations supplies, mobilize and equipment set
- *o* Design & Review - A/E fees, consultants, testing, models, & data processing
- *o* Direct Labor - associated with mobilization, installation, and administration
- *o* Other - utility connection, legal fees, appraisal, and waste management equipment

Operations & Maintenance Costs - Recurring Expense

- *o* Materials - parts, supplies, process chemicals, and incidental tools
- *o* Direct labor - for operating equipment, supervision, maintenance & contract labor
- *o* Operating Overhead - payroll charges, shipping, transportation, insurance, & rentals
- *o* Utilities - fuels, water, energy, and sewerage
- *o* General Administration - indirect labor, interest, travel, communications, & marketing

(2) Indirect Costs*

Compliance Costs

- *o* Notification - based on directives to comply in time or frequency
- *o* Reporting - preparedness, medical surveillance, loaded wage rates
- *o* Monitoring/Testing - planning, studies, modeling, inspections
- *o* Recordkeeping - maintain files associated with regulatory activities
- *o* Manifesting - listing/labeling hazardous process materials
- *o* Others - recovery cost, maintenance contracts, waste disposal, safety, & closure

Insurance

- *o* Worker - incremental cost of higher premiums paid due to risk
- *o* Third Party - incremental cost of higher premiums paid due to risk

On-Site Waste Management

- *o* Waste Management - collection, storage, transportation, sampling, and disposal
- *o* Non-recovered Materials - incremental cost of lost marketable by-product

* These cost have been traditionally hidden in the sense that they have been considered a burden to overhead in the past, or treated as externalities in conventional from the project cost analysis. These costs are in reality part of the production process or the product (EPA, 1989: p.3-6).

(3) Liability Costs

Fines & Penalties
Remediation

- *o* Air - costs associated with liability under federal, state, and local regulations
- *o* Soil - costs associated with liability under federal, state, and local regulations
- *o* Water - (groundwater & surface water) costs associated with liability under regulations

Containment

- *o* Waste Disposal - eminent liability associated with surface sealing

Legal Fees

- *o* Personal Injury - third party lawsuits seeking compensation for bodily injury
- *o* Economic Loss - claims and internal incremental costs of production loss
- *o* Real Property Damage - third party claims for loss of property value
- *o* Natural Resource Damage - claims seeking compensation for natural resource damage

(4) Intangible Costs

Qualifiers & Irreducibles

- *o* decreased readiness from distressed product quality
- *o* decreased standards due to poor image- organization & product-marketing
- *o* increased health maintenance costs due to current exposure
- *o* decreased efficiency from poor employee productivity/relations
- *o* increased production costs due to waste management decisions
- *o* increased operations & maintenance of facilities costs due to inefficient processes

When capturing the total costs of the alternative technologies we should attempt to quantify as many direct cost items as necessary; there will be little debate arising from the inclusion of the initial acquisitions cost or recurring O&M expense. The more controversial aspects of the analysis will stem from the attempt to quantify compliance or liability costs.

Although these issues require the expenditure of *real* dollars, the line items would also have to be extracted from the accounting system, which would normally require "breaking out" costs that are normally considered operational overhead. This paradox will make the capture of these costs difficult. For instance, liability costs are generated by fines and penalties usually levied for non-compliance (White et al., 1991: p.11). Most organizations are not well equipped to cost account for non-compliance, but business as usual can lead to the excessive costs due to this shortcoming. Moreover, legal awards or settlement costs can stem from remedial action and accidents causing personal injury or property damage. Superfund holds corporations financially responsible for environmental damage caused by previous waste disposal and management practices. Liability costs are difficult to estimate and predict their entry point into the life-cycle, so an emerging technology that effectively reduces the liability should be accounted for in a way that illustrates the potential cost savings. When future liability costs are included in the evaluation, the cost analyst introduces non-traditional uncertainties to decision makers (Peer & Beetle, 1990). For this reason, these potential costs are frequently omitted from the process, and if considered in the project analysis at all, management normally exercises caution in assigning a dollar value estimate to liability costs. The approach is too conservative and not realistic because these are real costs.

The initial research and development costs are generally not included, unless such costs are included for the base technology. It is crucial to perform cost analysis in an equitable atmosphere, in regard to both initial expenditures and future benefits. Benefits are more often considered to be costs avoided (Tarditi et al.,1995: p.18). In the case of some source reduction activities, there is no research and development, but decision costs are always present. Changing the way a filter is washed or the way a truck is unloaded might have no capital costs whatsoever, but have future benefits to the organization that can only be calculated by capturing the total costs of previous methods.

The less tangible costs and benefits are difficult to predict and estimate, but benefits or cost reductions of one system over another should be part of the evaluation of the assemblies because these benefits are real, even if there not easy to quantify (U.S. Army COE, 1995: p.8) The goal is to assign monetary value to environmental line items that have so far escaped analysis. We are attempting to include a range of value for liability costs, so we should proceed cautiously, as not to over or under value the less tangible costs. Although qualitative analysis may be more appropriate and salable to in some cases, overstating qualifiers could be an impediment to promoting sound business decisions that take environmental factors into account, when selecting the alternative with the strongest economic benefit. In the final analysis, the

bottom line will be the focus of the decision and any qualitative scoring or ranking system, no matter how well constructed will be looked upon as supporting information. Managers will typically skip ahead to the final number.

In addition, costs should be captured in a way that reflects the manner in which they were incurred. It is important not to cross operating centers when allocating costs; the type and quantity of contaminant reduced per center is more useful data than collecting the total for administration, research and development, etc. If possible, we should attempt to keep costs, such as those attributable to compliance, out of the general overhead category and move as many line items as possible to the direct cost category. This will focus attention on the proper source of the cost and make comparison of technologies more valid (West, 1993).

Life Cycle Cost Estimating

The time value of money becomes an important issue for costs that span more than one year. Techniques for making the analysis equitable over time will be applied to environmental technology projects. If we expand the list of cost components, then, at times, it is also necessary to look at an elongated time line for realizing the benefits of the investment. For instance, some pollution prevention projects might take many years to document costs and savings. A conventional time horizon for industrial project financial analysis might be less than five years. Accepting this typical five-year timeline could undermine the expanded cost-benefit component approach listed above. The *reason* for considering an extended analysis period can be demonstrated by examining the life-cycle cost for conventional pump-and -treat remediation of a contaminated site, which is a process that will rather commonly exceed a 30-yr span (Pendergrass, 1991). The decision maker's willingness to work with an extended time period for analysis will depend on funding, size and structure of the organization, process lifetime, and finally, return on investment from competing projects (White et al, 1992-93: pp. 38).

Calculating the return on investment for new technology requires incorporation of long-term financial indicators in the decision-making process. Assessment tools must consider the time value of money, and positive and negative cash flows over the life of the project. The tools for developing the Life-Cycle Cost Estimate (LCCE) are accepted economic analysis standards: Internal Rate of Return (IRR), Net Present Worth (NPW), and Benefit /Cost Ratio (BCR). All three procedures can appropriately discount future cash flows. The ROI is closely related to these assessment tools, which solve the same equation for different variables, and precede the

calculation of the return on investment. The selection of the assessment tool depends on the type of analysis required; the nature of total costs captured, and the desired expression of the final result, should lead to the proper analysis technique (General Electric, 1987). Key differences may include a increase in the number of line item costs, or an increase in the time period of the life-cycle, when appropriate. The use of these assessment tools is valid and straight forward, as long as the costs were captured in a comprehensive manner. Other methods, such as simple payback method, do not take into account cash flow beyond the "break-even" point or the cost of capital, but are none-the-less overused by large organizations as an inappropriate substitute for a more in-depth return on investment (ROI) calculation.

The calculation of Net Present Worth (NPW) is based on a known, or more likely assumed, discount rate. The sum of the discounted cash flows is the NPW of the project. If the project is worth pursuing, then the NPW is positive; a project with a negative NPW should be rejected. In general, a project with the higher NPW should be chosen over a lower NPW, if other parameters are equal. The calculations will illustrate this present worth method as sensitive to the rate of discount. This is particularly evident when NPW is applied to longer-term initiatives with substantial cash flows in later years. For a "frontloaded" project with most cash flows occurring in early years, the NPW will not be lowered much by increasing the discount rate. The opposite is true for projects whose major cash flow comes later. This means that, when using this method, projects with a big payoff towards the end of their life-cycle could be presented by the calculations as a less than attractive investment. For this methodology, the NPW of life-cycle costs is the present worth of capital costs, plus the present worth of the annual operations and maintenance costs, plus the present worth of the indirect costs and liabilities, if known, for each year the project is operable (Macedo et al., 1978: p.295). If the technology warrants the inclusion, the present worth of the salvage value of the equipment is also considered. The following formula would be used for a ten year life cycle: $\boldsymbol{NPW} =$

$$\sum_{a=1}^{F} PW_a = PW_1 + PW_2 + PW_3 + \ldots.. PW_9 + PW_F + PW_{salvage}$$

The present worth of each year through ten plus the salvage value is calculated by $PW = F(1+i)^{-n}$ where $\boldsymbol{P}$ represents a present sum of money, $\boldsymbol{F}$ is a future sum of money, $\boldsymbol{i}$ equals the interest rate per interest period (normally one year), and $\boldsymbol{n}$ is the number of interest periods. It is often necessary to calculate present worth for use in techniques other than *Net* present worth, as we will see below. There are many sources for compound interest factors at different rates. For instance, if we wanted to calculate the present worth of a series of equal

annual cost that will occur for the next ten years and the discount rate is assumed to be 10%, then we would multiply one year's cost by (6.145), which is the value for (P/A, 10%, 10 years. This factor can be found on page 630, of Newman & Johnson's "*Engineering Economic Analysis,*" Fifth Edition, 1995. This factor eliminates the need to individually calculate each annual cost in the series using the formula above. Once we have established a fundamental understanding of the methods, a spreadsheet could be constructed to automate any of these calculations and perform sensitivity analysis for various circumstances. Since one of are goals is to increase the understanding of the process, it is hardly a waste of time to look at what might be *behind* the use or construction of such a spreadsheet.

Using the internal rate of return (IRR) method, the discount rate that equates the present worth of cash inflow to present worth of expected project costs is calculated. A new technology would be worth using when the calculated IRR is greater than the cost of capital to finance its implementation. For the IRR, the net present worth is set to zero; the discount rate ***i***, is calculated. In a situation where a minimum interest rate of return, sometimes called a "hurdle rate" has been established, and where several competing projects require analysis, the project having the greatest IRR would theoretically be selected. To calculate the internal rate of return, we must convert the various consequences of the investment into a cash flow. Then we will solve for the unknown value of ***i***, which is the interest rate of return. There are five forms of the cash applicable to return on investment. **PW of benefits - PW of costs = 0, EUAB - EUAC = 0, PW of benefits / PW of costs = 1, NPW = 0, and PW of costs = PW of benefits.** These five equations represent the same concept in different forms (Newman & Johnson, 1995: p.165). The IRR method can relate costs and benefits with rate of return ***i*** as the only unknown. The calculation of IRR would be used in situations where decision makers wanted the final result expressed in percentage (%) form. The difficulty in solving for an interest rate is that there is no convenient direct method of solution. We solve the equations by trial and error, until one of the five conditions above is satisfied. A spreadsheet or software designed for this iterative process would be particularly useful for calculating the IRR.

The IRR is defined as the equivalent rate of return at which competing alternatives are equally attractive, For some firms , this is considered to be the ROI. Decision makers typically know the "hurdle rate" for investments; if the rate of return is above the hurdle rate, then the investment is acceptable. There are two main considerations that have a bearing on what interest rate to use in governmental investment studies. One obvious factor is the interest rate on borrowed capital, and the other is the sometimes overlooked opportunity cost of capital to the

governmental agency and to the taxpayers. At present, the federal government uses a 10% standard hurdle rate, according to the Office of Management and Budget (OMB, 1992). OMB sets the guidelines for projects sponsored by the government. This factor is based on a minimum rate calculated by subtracting the projected average annual rate of inflation from the nominal annual interest rate for Treasury notes and bonds. Currently, the factor for constant dollar analysis of ten year projects is would be about 4 percent (Ewer, 1992: p.71). Rates established by private businesses are dependent are relevant to this study only so far as they effect subcontract prices.

The Benefit/Cost Ratio (BCR) is sometimes called the Profitability Index. The BCR amounts to the present value of cash flow in (benefits) over the present value of cash flow out (costs). This illustrates the present worth of dollar value benefits per dollar spent or the relative profitability of the project. Projects with the highest ratio greater than one should be pursued. In governmental projects there may be difficulties deciding whether to classify various consequences as items for the numerator or for the denominator. An alternate computation for public funded projects is to consider user costs a *dis*benefit and to subtract them in the numerator rather than adding them in the denominator (Newman & Johnson, 1995: p.426).The reason for this suggested alteration to the common benefits/costs equation can be illustrated by using the example of a government project with the following consequences:

- *Initial cost of project to be paid by government is* \$1,000,000
- *Present Worth of future maintenance to be paid by government is* \$423,000
- *Present Worth of Benefits to the public is* \$3,336,000
- *Present Worth of additional public user costs is* \$617,000

If we put the benefits in the numerator and all the costs in the denominator it yields:
BCR = \$3,336,000 / \$1,000,000 + \$423,000 + 617,000 = \$3,336,000 / \$2,040,000 = **1. 64**

Using the alternate calculation below to consider user cost as a disbenefit, since the people receiving the benefits may pay none of the costs directly, would compute:

$$BCR = \frac{\textbf{Public benefits - Public costs - Maintenance costs}}{\textbf{Governmental costs}}$$

BCR = \$3,336,000 - \$617,000 - \$423,000 / \$1,000,000 = \$2,296,000 / \$1,000,000 = **2.30** <——

We should note that, while this will yield a higher benefit - cost ratio than may be conventionally calculated, the NPW does not change.

NPW = PW of benefits - PW of costs = \$3,336,000 - \$617,000 - \$423,000 - \$1,000,000
= \$1,296,000

Again, these are rather commonly used economic analysis techniques, but there were some new line item costs identified earlier that, in conjunction with the often extended analysis period, separate the analysis of emerging environmental technologies from typical industrial projects requiring similar analysis (DeGarmo et al., 1993; Newman & Johnson, 1995; Macedo et al., 1978). The key to assessing the economic viability of investment in new technology is to open an organization's accounting system, so that it can be used to track and allocate environmental costs to the process responsible for creating them. If this is done properly, the cost accounting system can provide relevant cost data and the time/rate constraints for analysis of the true life-cycle costs and operating budget. We should pursue the assessment of life-cycle cost in enough detail to allow for calculation of a reliable return on investment.

Return on Investment

The potential profitability of investing in an emerging environmental technology could be expressed in a variety of calculations described as return on investment. The IRR is defined as the equivalent rate of return at which competing alternatives are equally attractive--expressed a percentage rate. The ROI is defined in this report as the actual dollar benefit of an investment. This is a more straightforward expression than profitability percentages, especially when developing budgets within government agencies. For these organizations, reaching a certain profit-making percentage is not their primary mission. Furthermore, a relatively high *incremental* rate of return, of say--40%, placed in a final report, could mislead or encourage managers to believe that the entire project has a 40% *internal* rate of return.

Return on Investment (ROI) is a criterion for judging the most efficient way of solving a given problem. The selected method should be the most efficient, that is, the least expensive way of meeting the performance objectives of the project. No matter what other decision making rule has been applied, we calculate the expected savings, from employing the new technology, relative to the existing alternatives. Calculating return on investment requires understanding the scope of the environmental problem and an estimate of the number of times the new technology would actually be used to solve the problem. The return on investment would ultimately equal the

discounted cost savings minus the discounted cost of initial investment over a period of time. This final step is the culmination of our efforts to calculate the total life-cycle costs associated with alternative approaches to environmental solutions and should serve as a useful decision making tool, when presented to management with proper qualifications.

Summary of Methodology

The steps involved in evaluating return on investment for new technologies are to (1) **specify the cost problem**, using established goals and objectives and relevant data to define the alternatives, (2) **develop the decision criteria model**, so the performance of the competing technologies can be characterized within the context of application, and the predominate cost factors will be exposed, (3) **capture the total costs** to implement the alternatives, including direct, indirect, liability, and qualify the less tangible costs, (3) **estimate the life-cycle costs** associated with costs and benefits occurring at different times, while applying present worth methods with discount rates and life-cycles appropriate to the specific project, (4) calculate **return on investment** expressed as an actual dollar amount benefit of implementing the emerging technology. Collectively, this process should provide structure and promote consistency in evaluating investments in new technology.

Calculating Return On Investment
For Emerging Technologies

SUMMARY OF METHODOLOGY

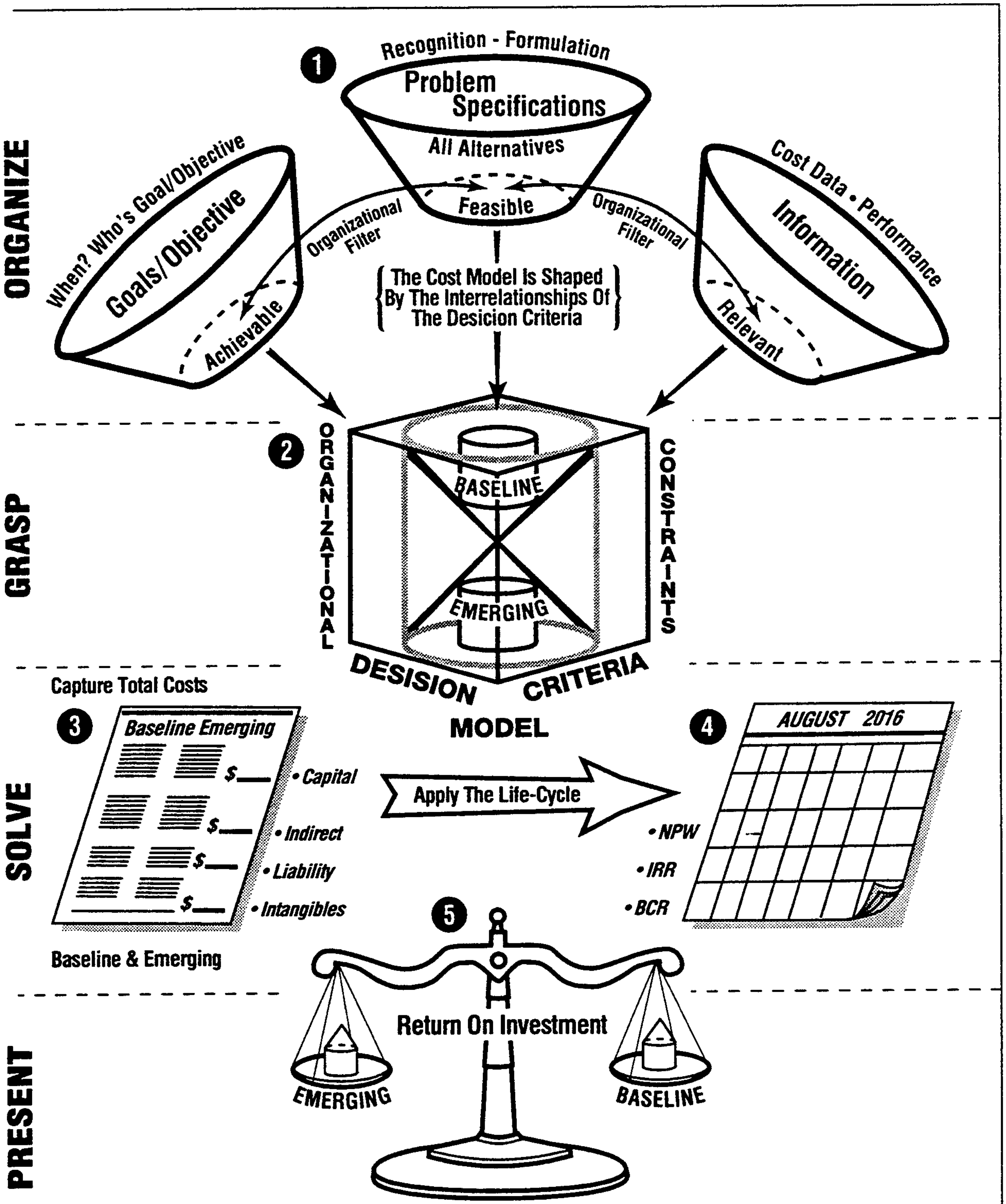


EXHIBIT A

JAVA-BASED APPLICATION OF THE MODEL-VIEW-CONTROLLER FRAMEWORK IN DEVELOPING INTERFACES TO INTERACTIVE SIMULATIONS

S. Narayanan
Assistant Professor
Department of Biomedical and Human Factors Engineering;
and
Nicole L. Schneider
Graduate Teaching Assistant
Department of Biomedical and Human Factors Engineering

207 Russ Engineering Center
College of Engineering and Computer Science
Wright State University
Dayton, OH 45435
Tel: +1 513 873 5071
Fax: +1 513 873 5009
Email: snarayan@cs.wright.edu

Final Report for:
Graduate Student Research Program
Armstrong Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, DC

and

Armstrong Laboratory

September 1996

JAVA-BASED APPLICATION OF THE MODEL-VIEW-CONTROLLER FRAMEWORK IN DEVELOPING INTERFACES TO INTERACTIVE SIMULATIONS

S. Narayanan and Nicole L. Schneider
Assistant Professor
Department of Biomedical and Human Factors Engineering
Wright State University
Graduate Teaching Assistant
Department of Biomedical and Human Factors Engineering
Wright State University

Abstract

Interfaces to simulations serve to portray the dynamic behavior of the modeled system. In visual interactive simulations, user interfaces allow an analyst to also interact actively with the executing simulation. Traditionally, the software to display the simulation model and to facilitate user interaction are embedded in the simulation model. Such an integration makes it difficult to maintain large simulation programs and pose limitations in the development of multiple interfaces to a simulation model. This article presents a Java-Based Architecture for Developing Interactive Simulations (JADIS). JADIS applies the Model-View-Controller paradigm to the development of interactive simulations. In JADIS, the simulation model and multiple interfaces to them are separate processes that execute concurrently on distributed machines. JADIS integrates concepts from object-oriented programming, concurrent, distributed processing, and graphical user interface design in developing visual interactive simulations. This article describes the JADIS architecture and presents application of JADIS to the airbase logistics modeling domain.

JAVA-BASED APPLICATION OF THE MODEL-VIEW-CONTROLLER FRAMEWORK IN DEVELOPING INTERFACES TO INTERACTIVE SIMULATIONS

S. Narayanan and Nicole L. Schneider

Introduction

Discrete-event simulations offer flexible and powerful means of observing and analyzing complex, dynamic systems. The fundamental objective in simulation is to use the software abstractions provided by the language to represent the behavior of system entities over time. In traditional discrete-event simulations, the analyst is primarily a passive observer of the simulation program execution. With the increase in computing power and graphical user interfaces, there is an increasing interest in the area of visual interactive simulations (VIS)[Bell & O'Keefe, 1987; Bell, 1991; Hurrion, 1980; Lyu & Gunasekaran, 1993; McGregor & Randhawa, 1994].

In VIS, interfaces serve to not only display the state of the simulated system, but also to allow an analyst to interact with the executing simulation. The simulation executes in real-time or scaled time. The analyst can modify the parameters of the simulation, alter the dynamics of the simulated system, and can pause/restart the simulation. The VIS approach offers several potential advantages. First, it allows the user to make complex decisions. For example, Hurrion and Secker (1978) found that the rules used by human schedulers in job shop scheduling were difficult to encapsulate in simulation models. VIS offered a viable alternative. Second, VIS are useful in studying the effectiveness of real-time, human decision making in complex systems. Dunkler, et al. (1988), for example, used an interactive simulation of a flexible manufacturing system and compared the effectiveness of various automatic scheduling strategies with that of human scheduling in expediting parts through the system. Third, the display of the simulated system in VIS can be visually appealing and can increase effective communication between a manager and the simulation analyst in model development (Bell, 1991; Bishop & Balci, 1990). Fourth, the dynamic visual representation in VIS can highlight logical inconsistencies in the model and can therefore be effective in model verification and validation. Finally, since the user of VIS actively participates in the execution of the simulation, there is potential for increased user confidence in applying the results of the simulation (Kirkpatrick & Bell, 1989).

There are several potential problems with the VIS approach (Bell & O'Keefe, 1987; Bishop & Balci, 1990; Paul, 1989). First, due to human interaction at various times during the execution of the simulation, the simulation experiments are hard to duplicate and are not amenable to traditional simulation output statistical analysis. Second, a user interacting with the simulation may observe a snapshot of the system and may prematurely conclude that the system will always exhibit the observed characteristics without the benefit of detailed analysis.

Third, in the VIS approach, design of the dynamic display tends to be an integral part of the simulation model development making the traditional simulation life cycle inadequate to describe the VIS approach (Hurrion & Secker, 1978).

Despite the problems outlined above, when applied appropriately, interactive simulations are useful in the analysis of complex, dynamic systems. They are necessary to analyze human interaction with complex systems and can be effective in enhancing user understanding of large, semi-structured problems through interaction with the simulation.

The major challenges in developing interactive simulations are problems associated with computer hardware and software (Bell & O'Keefe, 1987). Bell (1991) highlights the historic struggle of the early VIS development effort with advances in computer hardware. Early VIS systems including See-Why were developed for large main frames. Currently, personal computers and workstations have become the standard for systems development. Most VIS packages currently available are still hardware dependent and suffer from problems of portability.

Several early interactive simulation packages were developed in FORTRAN (e.g., FORSSIGHT). Developmental interest has moved towards C and recently towards object-oriented languages (e.g., ProfiSEE in Smalltalk-80 [Vaessen, 1989]). While object-oriented programming offers many advantages for simulation modeling in terms of modularity, software reuse, and natural mapping with real world entities (Narayanan et al., 1996), their application to developing interactive simulations has been only explored in a limited way (Bell, 1991). The software to display the simulation model and to facilitate user interaction are embedded in the simulation model. Such an integration makes it difficult to maintain large simulation programs and pose limitations in the development of multiple interfaces to a simulation model. Although it is acknowledged that the interface configuration and interaction specification are concurrent with the model specification, effective means to facilitate concurrent software development are lacking. As a result of the tight coupling of the simulation model with the interface, it is often difficult for concurrent development of the two phases.

This article presents a Java-Based Architecture for Developing Interactive Simulations (JADIS). JADIS applies the Model-View-Controller paradigm from Smalltalk to the development of interactive simulations (Goldberg, 1990). In JADIS, the simulation model and multiple interfaces to them are separate processes that execute concurrently on distributed machines. JADIS integrates object-oriented programming, concurrent, distributed processing, and graphical user interface design in developing visual interactive simulations.

The remainder of the article is organized as follows. First, we present background to visual interactive simulations and the model-view-controller framework. We then discuss the motivations for applying the MVC framework to developing interfaces to interactive simulations and present the JADIS architecture. We will then describe the application of JADIS to airbase logistics. We will discuss our approach in the context of existing efforts in the literature and conclude with a summary of contributions of this study and suggest recommendations for future work.

Background

Table 1 presents a comparison of interactive simulations with traditional discrete-event simulations and animated simulations along seven dimensions: nature of suitable problems, simulation development life cycle, time transition of simulation clock, nature of user interaction, role of the graphical interface, types of output analysis, and example of software packages for each category. Interactive simulations are well suited for large, semi-structured problems in which human interaction is an important consideration. Interactive simulation development is different from the traditional simulation life cycle as the specification of interaction and animation is concurrent with model specification. The simulation clock is updated either on a real-time basis or on a scaled time. The graphical interface in interactive simulations depict dynamic system states, highlight performance measures, and contain interface objects that accommodate command line inputs and other user interaction. Output analysis in interactive simulations primarily involves transient systems analysis.

Table 1. Comparison of Interactive Simulations with Traditional Discrete-Event Simulations and Animated Simulations.

Topic/Issue	**Traditional discrete-event simulations**	**Animated discrete-event simulations**	**Interactive simulations**
Nature of suitable problems	• Well-structured problems • Small to medium scale systems • Human interaction not a critical consideration	• Well-structured to semi-structured problems • Medium to large scale systems • Human interaction not a critical consideration or can be captured completely	• Semi-structured to unstructured problems • Small to medium scale systems • Human interaction is a critical consideration
Simulation development life cycle	• Traditional	• Traditional with animation specification following simulation model configuration	• Interface and interaction specification concurrent with simulation model specification
Time transition of simulation clock	• Discrete-event to another	• Discrete-event to another with scaled time for animation	• Discrete-event to another • Scaled time • Real time

User interaction	• None	• None or able to alter simulation experimental parameters	• Able to alter simulation experimental parameters • Able to alter simulation system dynamics
Graphical interface	• None	• Displays dynamic system states • Displays performance measures	• Displays dynamic system states • Displays performance measures • Interface objects accommodate command line input and other user interaction
Output Analysis	• Conventional steady state or terminating simulation analysis	• Conventional steady state or terminating simulation analysis	• Transient systems analysis • Human factors engineering analysis methods
Software	• Simulation languages (e.g., GPSS V) • Programming language (e.g., FORTRAN)	• Simulation packages with animation capability (e.g., SLAM/TESS, SIMAN/CINEMA) • Programming language (e.g., FORTRAN, C)	• Simulation packages (e.g., VISION, SEE-WHY, WITNESS) • Programming language (e.g., C, C++, Smalltalk)

The major challenge in developing visual interactive simulations is associated with computer hardware and software problems (Bell, 1991). There is need for computational architectures that can enable the development of interactive simulations hardware independent. Also, object-oriented languages offer tremendous promise and are obvious vehicles for VIS development (Bell and O'Keefe, 1997). Several packages such as Audition and ProfiSEE (Vaessen, 1989) have been developed using an object-oriented language. Existing packages, however, exploit the power of object-based programming and concepts such as model-view-controller framework in only a limited manner. Most architectures are also hardware dependent.

This article discusses a Java-based Architecture for Developing Interactive Simulations (JADIS). JADIS overcomes the hardware and software limitations of traditional VIS architectures outlined above. JADIS is developed using the Java programming language (Lemay & Perkins, 1996). The Java source code is compiled into byte code that can be read by an interpreter available on multiple platforms including personal computers, Macintoshes, and UNIX workstations. The software can be developed on any platform that contains the Java Development Kit (JDK). JDK is available on most operating systems. The byte code can then be moved to another platform and can be run successfully without altering a single line of code. Figure 1 illustrates the Java virtual machine environment.

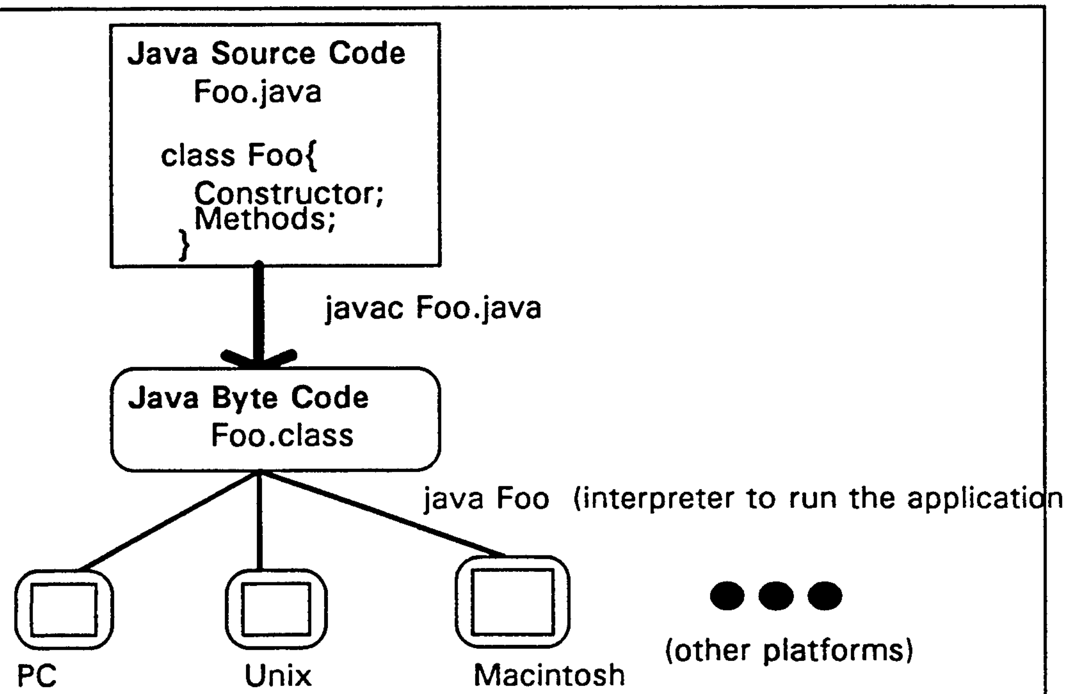


Figure 1. Illustration of the Java virtual machine environment.

Java is an object-oriented programming language whose syntax is similar to C++. Java supports encapsulation, inheritance, and polymorphism. It, however, does not have explicit pointers, doesn't support multiple inheritance, and doesn't feature operator overloading. Java's popularity is initially because of applets, which are written in Java and can be embedded on home pages on the world wide web. Applets add animation and interaction to web pages and can be viewed using a browser such as Netscape 2.0 or higher. Java can also be used as a regular programming language where application can stand alone without being embedded as applets. The Java language comes with various packages (similar to libraries) for general data structures, applets, file input/output, and also for graphical user interfaces. Java is multithreaded and hence particularly suitable for distributed computing as it easily copes with TCP/IP protocols. Java can thus be used for both creating simulations as well as for creating interfaces to simulations. The Java language also features a utility called javadoc which enables automatic hypertext generation of software documentation. Users can add specialized comments in the source code which can be easily processed by javadoc in generation of source code documentation.

JADIS applies the Model-View-Controller (MVC) paradigm from Smalltalk to the development of interactive simulations. We first describe the MVC paradigm before describing JADIS.

MVC is a paradigm for developing graphical user interfaces in a modular manner (Gobbetti and Turner, 1991, Goldberg, 1990; Krasner and Pope, 1988). In MVC, any interactive program is conceptually divided into three areas: (1) the Model, which contains representation of the application domain, (2) the View, which contains specification of the display, and (3) the Controller, which contains specification of user interactions with the underlying model. In the context of interactive simulations, the model refers to the simulation model, the view corresponds to how the dynamics of the simulation are displayed to the user, and controller refers to the processing of user commands input to the simulation model through the display.

The MVC framework provides several potential advantages in developing interfaces to interactive simulations. First, due to the separation of the model from the view, the simulation model development can take place concurrently with the specification of the interface. The simulation developer can focus on the model development and leave the responsibility of the interface design to a graphical user interface (GUI) designer. Second, multiple views to the same simulation can be developed. The end user can then plug different displays or pieces of code into the simulation. The simulation model can be arranged to suit the needs of individual modelers without requiring programmers to constantly create entirely new code. Thus, the productivity of software development is enhanced. The reuse of existing designs and refinements potentially also leads to stable applications with a consistent style.

MVC is an improvement over previous approaches to developing interactive simulations. Simulation modelers need no longer be experts at implementing simulation models as well as be able to design and implement graphical displays. GUI experts can create display modules which can be stored in graphical libraries for the simulation analysts to use in customizing the simulation view. With MVC, many users can access multiple, simultaneous views of the same simulation model.

The JADIS Architecture

Figure 2 presents a schematic of the JADIS architecture. There are three primary modules in JADIS: (1) simulation module comprising the simulation infrastructure including clock, random number generators, various statistical distributions, and event calendar; and domain-specific classes consisting of general queuing utilities and system-related classes (for example in air base logistics simulation classes include air base, hangar, and air craft); (2) interface module consisting of general display classes (e.g., list boxes) and domain-specific display elements (e.g., air craft), and (3) inter-process communication module consisting of sockets and the ability of the simulation to broadcast state changes to the views and the ability of the interface process to send commands to the simulation and receive messages from the simulation. The simulation process instantiated from the simulation module and the interface process instantiated from the interface module can run concurrently on the same machine or on different

platforms. Users can input commands to the simulation through the command line or through the interface. Messages travel between the simulation and the interface processes in the form of text strings.

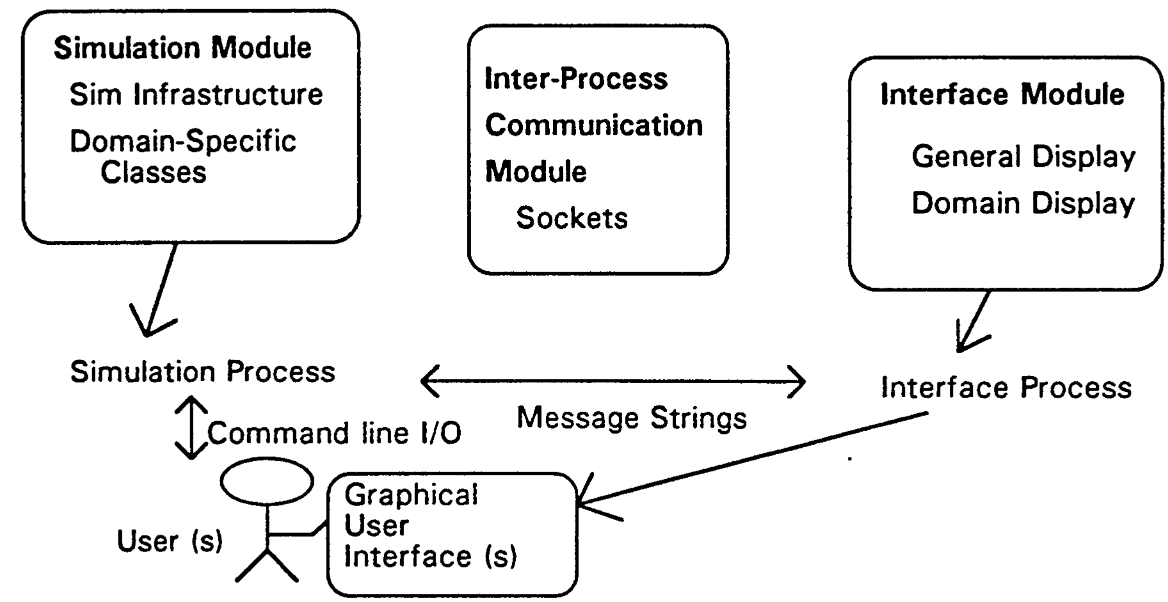


Figure 2. Schematic of the JADIS Architecture.

The JADIS architecture alters the simulation model development process. In using JADIS, the analyst defines the elements of the simulation model. The analyst then defines the interface and specifies the communication messages. The simulation model and views are thus concurrently specified. The analyst creates instances or subclasses as appropriate and interconnects the simulation and interface processes. The software is then tested, model is verified, and hypertext source code documentation is automatically generated through the `javadoc` utility in Java. The JADIS architecture has been applied in the interactive simulation of airbase logistics. The application is discussed in the next section.

<u>Application to Airbase Logistics</u>

The domain of airbase logistics is large and complex. It involves logistics processes that support aircraft sortie generation at operational airbases. Airbase logistics involves aircraft maintenance, parts supply, and munitions loading (Popken, 1992). Models of logistics processes are useful in analysis for aircraft acquisition planning, maintenance manpower allocation, and theater-level supply redistribution. Popken (1992) discusses that a synergistic combination of object-oriented programming, databases, and graphical user interfaces provide

significant enhancements to simulation modeling capabilities. An aircraft maintenance problem provided an excellent testbed for demonstrating the JADIS architecture.

This section is organized as follows. First, we outline the features of the system and the assumptions made. We then describe the principles in designing the simulation model and the interface. We discuss the various classes and highlight the salient class hierarchies, and detail the capabilities of the system. We also describe the lessons learned during the model development process.

The model used is designed to capture maintenance operations at an airbase. There are aircraft of different kinds with varying configurations and capabilities. An aircraft is comprised of several subsystems. Sorties are generated by different methods (e.g., random generation). Each sortie specifies the number of aircraft required, the type of each aircraft, and the details of the mission. While the aircraft is in operation one or more of its subsystems may fail. When a subsystem of an aircraft fails, it is sent to the maintenance facility for repairs. The maintenance facility includes a hangar, different types of *test* equipment, spare parts, and personnel. Various performance measures in this system include maintenance cost, sorties completed, sorties aborted, hangar utilization, and personnel utilization. We developed classes to represent the entities in this system and to specify their interaction (Carrico & Clark, 1995; Carrico et al., 1995).

The specific implementation made some simplifying assumptions. First, the model was at the airbase level and not at the theater level. Second, the maintenance resources were always available. Third, subsystems featured a single failure. Each subsystem failure identified the unique maintenance actions required. Sorties are generated randomly, in a pre-set mode, or in the fly-when-ready mode. The simulation duration was two weeks, with sortie generation occurring 16 hours each day, seven days a week.

Several design principles were applied both in the simulation model development and in the interface design. First, we exploited the capability of natural mapping and modularity features of object-oriented programming. Through object-oriented programming it is possible to develop software abstractions that have a direct correspondence with real world objects (Narayanan et al., 1996). Objects can also be reused through inheritance. Second, in the simulation model, physical objects (e.g., aircraft, subsystem, and hangar) were distinguished from decision making objects (e.g., scheduler, resource manager) and information storage objects (e.g., resource statistics). The advantage of making such a distinction is to allow different decision making strategies to be evaluated using the simulation model, where only the decision making entities need to be changed. Third, the interactions between objects were limited thereby enhancing the plug-in capability of the architecture.

On the interface development process, information was presented hierarchically. The analyst can get to see the overall dynamics of the simulated system and if desired could then look at details of the individual components. Second, the interface accommodates users in multiple modes where the users interest may be in systems analysis or simply in visualizing the maintenance processes. Third, interfaces feature a standard look and feel as that of Motif. It also features balloon help as in Microsoft Windows. Finally, the components of the architecture for interface development can be easily assembled for rapidly prototyping graphical user interfaces for a class of similar problems.

The application developed consists of two distinct processes: (1) simulation and (2) interface. These processes can either run on the same machine or different machines. All simulation classes are descendants of `SimBase`. Figure 3 highlights salient top-level classes in the simulation model. Two major subclasses of SimBase are `ActiveSimulationObject` and `InformationStore`. `DecisionMaker` and `Physical` are two subclasses of `ActiveSimulationObject`. Subclasses of `InformationStore` include `ResourceStatistics` and `MaintenanceInfo`. Similarly, `Equipment`, `Personnel`, and `Hangar` are all subclasses of `Resource`. DecisionMakers include `ResourceManager` and `Coordinator`. All of these classes have a natural mapping to real world entities.

On the interface process, major classes include `Interface` which initiates threads for receiving input from simulation and for setting up the animation, class `Animation` (inherited from Java's `Frame` class) which sets up displays by instantiating the domain specific display objects such as hangar, aircraft, and runway, class `CommandEntry` which facilitates user interaction in a command window, class `Dynamics` which displays visualization of the processes involved in airbase maintenance, class `EventList` which displays a log of events as they occur on the simulation side, classes `GraphSS` and `Graphs` to display graphs of various performance measures, and classes for menus, push button, etc. overloaded from Java's abstract windowing toolkit. `Animation` has a `processEvent` method which in turn invokes `processEvent` of displayed objects. The knowledge of how displayed objects update to the events in simulation are encapsulated in the class representation of the displayed objects.

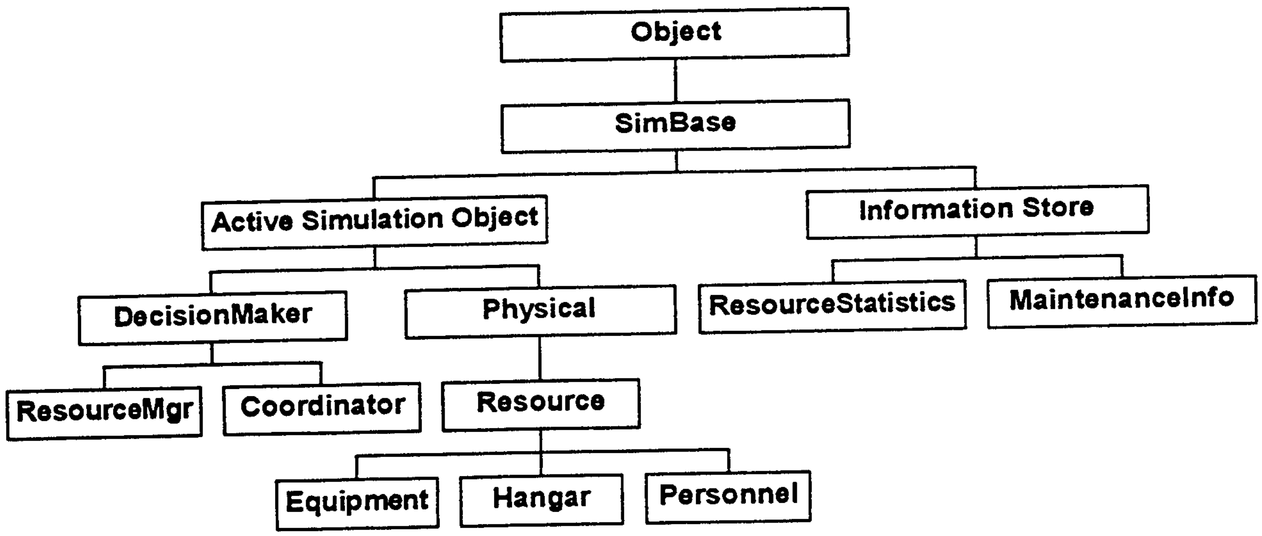


Figure 3. Illustrative hierarchy of simulation classes to model aircraft maintenance.

There were a total of 49 Java classes implemented for the simulation model of this application. There were 15 additional classes used for the interface. The size of classes ranged from 20 to over 800 lines totalling approximately 21,000 lines of code. Screen output of the main interface is included in the appendix. The javadoc capability of Java was used in automatically generating the source code documentation with hypertext links. Javadoc turns comments to code into hypertext markup language code with links to the parent class and overridden methods. The URL for the source code documentation of the application is `http://isis.cs.wright.edu:1947/fx99html/`. A sample javadoc generated file is included in the appendix for illustrative purposes.

Discussion

This report presented JADIS, a Java-based architecture for developing interactive simulations. The JADIS architecture integrated concepts from object-oriented programming, distributed computing, and graphical user interfaces to interactive simulations. Since the architecture is implemented in Java, it is hardware independent. The JADIS architecture is an instantiation of the Model-View-Controller framework in interactive simulations.

JADIS, however, goes beyond the traditional implementation of the MVC framework. While the MVC framework provides a powerful metaphor for developing interactive simulations, practical implementations often lead to complicated, unwieldy class inheritance structures (Krasner & Pope, 1988; Shan, 1990). In JADIS, the model and views are completely separated. The inheritance structure maps well to the real world objects. The semantics of how displayed objects are updated is encapsulated well within the class representation. Also, in the traditional

MVC, a model broadcasts that its status has changed, all views and controllers tied to that model are required to query it to discover exactly what the change is before they can update themselves. JADIS reduces redundancy in message passing by having the simulation model broadcast that its state has changed and the details of the state change to the views. Finally, in contrast to the polling protocol applied in the traditional MVC, JADIS applies an process-based, event-driven protocol which maps well with simulations where behavior is represented as events occurring at different time units.

The architecture also facilitates the instantiation of interactive simulations. This capability goes well beyond animating discrete-event simulations such as seen in MicroSaint or CINEMA for SIMAN. In JADIS simulations, users can alter the parameters of the simulation and also modify the system dynamics. For example, users in the airbase logistics simulation can alter the parameters of the maintenance resources at run time and also alter the sortie generation discipline. Real-time human decision making can therefore be readily studied using JADIS simulations. The ability to run the simulation and interface on multiple machines concurrently is also a powerful capability in harnessing the power of distributed computing.

The architecture currently has a few limitations. First, it has not yet been tested for handling simultaneous user input from multiple views of the same simulation model. Second, the simulation and the interface are implemented as Java applications rather than Java applets. Therefore, they can not directly be viewed using an internet browser. Third, when the number of messages between the simulation and the interface becomes very high, it slows down the machine. This limitation can be overcome by incorporating capabilities to filter needless messages appropriately. The airbase modeling application has some limitations as well. First, the model was developed at the airbase level and not at the theater level. Second, human interaction was limited to altering clock speeds, viewing performance measures, altering maintenance data files, and scheduling disciplines of sorties dynamically. Third, the maintenance behaviors were also simplified to feature single failure subsystem and also having adequate spares and other maintenance resources.

Future research extensions include incorporating the capability to run the simulation and view it on internet, extending the scope of the application to include modeling at the theater level, incorporating additional human interaction, enhancing the visualization capabilities in the system, and empirically evaluating the efficacy of interfaces tailored to users. The ultimate goal is to have a high-fidelity computational representation of airbase logistics so as to support logistical decision making through computer-based tools. Integrating interactive optimization capabilities to the descriptive simulation modeling architecture is another promising avenue of research.

Conclusions

Advances in software offer unique opportunities in enhancing simulation modeling capabilities. Interactive simulation is a useful methodology for systems analysis of large, complex systems. Traditionally, the development of interactive systems have been plagued with software and hardware problems. We have applied Java programming language and integrated concepts from object-oriented programming, model-view-controller, distributed computing to develop a JADIS, a Java-Based Architecture for Developing Interactive Simulations.

The JADIS architecture can run on any platform which supports the Java Development Kit. Such systems include PCs running Windows 95 or Windows NT, Macintoshes, and UNIX workstations running Solaris operating system. In the current implementation of JADIS, the simulation and interface are implemented as Java applications. Once they are converted to be designed as applets, then the entire architecture can be run using an internet browser such as Netscape on any machine.

The JADIS architecture was evaluated in the context of an aircraft maintenance problem in airbase logistics. The classes in JADIS for this application were based on a set of principles to enhance reuse, exploit natural mapping, and rapidly test different decision making strategies. The JADIS application for air craft maintenance is an interactive simulation accommodating active human interaction and has visualization capabilities.

Java was found to be a powerful language. The ability to readily move code between multiple platforms is a powerful feature. The large number of built-in classes in the language enhanced reuse. Java's capability to not only be applicable for implementing simulations but also its use in graphical user interface design made it very powerful to developing interactive simulations. Finally, the javadoc capability in Java to quickly generate hypertext source code documentation was a valuable feature.

The findings of this research contribute to the area of interactive simulations. The JADIS architecture offers a solution to the hardware and software problems encountered in interactive simulation development. Through application of the Model-View-Controller framework, simulation model development and interface design can take place concurrently thereby potentially reducing the simulation development lifecycle cost. The architecture also facilitates the study of human interaction with complex systems and the effectiveness of tailored views to interactive simulations. The airbase logistics problem studied in evaluating the JADIS architecture appears to be a ripe application area for implementing interactive simulations.

Acknowledgments

The research reported in this paper was sponsored by the United States Air Force, Logistics Research Division, Armstrong Laboratory, Wright-Patterson AFB, under contract with Research and Development Lab. We thank Capt. Todd Carrico and 2Lt. John DiPasquale at Armstrong Laboratory for their support and encouragement during this work.

References

Bell, P.C. & O'Keefe, R. M. (1987). Visual interactive simulation - History, recent developments, and major issues. *Simulation,* 49(3):109-116.

Bishop, J. L. & Balci, O. (1990). General purpose visual simulation system. *Proceedings of the 1990 Winter Simulation Conference.* 504-512.

Carrico, T., Clark, P. K., Shute, N. J., and Zahn, E. A. (1995). Integrated Model Development Environment (IMDE) multi-function aerospace support system. Technical report, *AL/HR-TR-1995-0186,* Armstrong Laboratory, WPAFB, Dayton, OH.

Carrico, T. & Clark, P. K. (1995). IMDE support for Air Force logistics. Technical report, *AL/HR-TR-1995-0187,* Armstrong Laboratory, WPAFB Dayton, OH.

Dunkler, O., Mitchell, C. M., Govindaraj, T., & Ammons, J. C. (1988). The effectiveness of supervisory control strategies in scheduling flexible manufacturing systems. *IEEE Transactions on Systems, Man, and Cybernetics,* SMC-18, 223-237.

Gobbetti, E. & Turner, R. (1991). Object-oriented design of dynamic graphics applications. *In New Trends in Animation and Visualization,* N. M. Thalmann and D. Thalmann Eds., Wiley Professional Computing, Sussex, England, 43-58.

Goldberg, A. (1990). Information models, views, and controllers. *Dr. Dobb's Journal,* July, 54-61.

Hurrion, R.D. (1980). An interactive visual simulation system for industrial management. *European Journal of Operational Research,* 5(2): 86-94.

Kirkpatrick, P.F. & Bell, P.C. (1989). Visual interactive modeling in industry: Results from a survey of visual interactive model builders. *Interfaces,* 19, 71-79.

Krasner, G. E. & Pope, S. T. (1988). A cookbook for using the model-view controller user interface paradigm in Smalltalk 80. *Journal of Object Oriented Programming,* August/September, 26-49.

Lemay, L. & Perkins, C. L. (1996). *Teach yourself Java in 21 days.* Sams.Net Publishing.

Lyu, J. & Gunasekaran, A. (1993). Developing a visual interactive simulation model for flexible manufacturing systems. *International Journal of Operations and Production Management,* 13(6): 59-67.

Narayanan, S., Bodner, D. A., Sreekanth, U., Govindaraj, T., McGinnis, L. F., & Mitchell, C. M. (1996). Research in object-oriented manufacturing simulations: An assessment of the state of the art. Submitted to *IIE Transactions*.

Paul, R. J. (1989). Visual simulation: Seeing is believing. In *Impacts of Recent Computer Advances on Operations Research*. R. Sharda, B. L. Golden, E. Wasil, O. Balci, and W. Stewart, Eds. Elsevier Science Publishing, New York, NY, 422-432.

Mathewson, S. C. (1984). The application of program generator software and its extensions to discrete event simulation modeling. *IIE Transactions*. 16(1): 3-17.

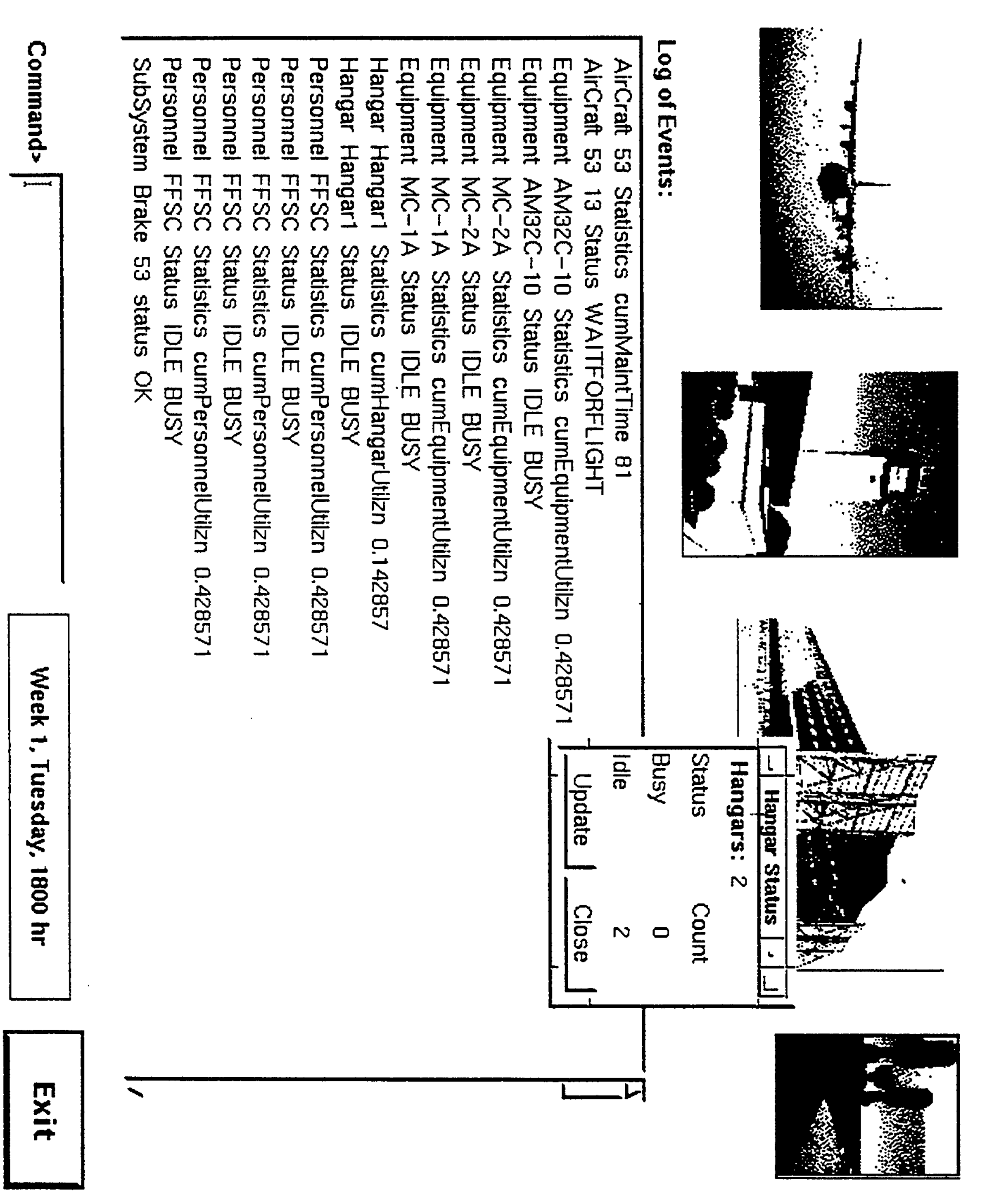
McGregor, D. R. & Randhawa S. U. (1994). ENTS: An interactive object-oriented system for discrete simulation modeling. *Journal of Object-Oriented Programming*, January, 21-29.

Popken, D. A. (1992). An object-oriented simulation environment for airbase logistics. *Simulation*. 59(5): 328-338.

Rooks, M. (1993). A user-centered paradigm for interactive simulation. *Simulation*, 60(3): 168-177.

Shan, Y-P. (1990). MoDE: A UIMS for Smalltalk. University of North Carolina, Chapel Hill, *TR90-017*, DTIC, April.

Appendix A: Screen output of the main interface.



Appendix B: Javadoc output for class `Rnormal`.

Class Rnormal

```
java.lang.Object
   |
   +----SimBase
           |
           +----Distribution
                   |
                   +----Rnormal
```

public class **Rnormal**
extends Distribution

The **Rnormal Class** creates a Normal Distribution with a mean of zero and a variance of one (i.e. N(0,1)).

Version:
July 16, 1996

Author:
S. Narayanan

Constructor Index

- **Rnormal**(long)
  The overloaded *Rnormal* Constructor calls the Distribution's constructor to set the seed

Method Index

- **getNextRnormal**()
  The *getNextRandom* Method gets next random number of the standard normal distribution.

Constructors

- **Rnormal**

```
public Rnormal(long firstSeed)
```

The overloaded *Rnormal* Constructor calls the Distribution's constructor to set the seed

Methods

• getNextRnormal

```
public double getNextRnormal()
```

The *getNextRandom* Method gets next random number of the standard normal distribution. It gives the n(0,1) deviate by composition method of ahrens and dieter (see Brately, Fox, and Schrage, pg. 318).

Examing Alternate Entry Points in a Problem Using Fuzzy Cognitive Maps

Karl Perusich,Ph.D.
Associate Professor
Department of Electrical Engineering Technology

Purdue University
1700 Mishawaka Ave.
South Bend, IN 46634

Final Report for:
Summer Faculty Research Program
Armstrong Lab

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, DC
and
Armstrong Lab

September 1996

Examing Alternate Entry Points in a Problem Using Fuzzy Cognitive Maps

Karl Perusich, Ph.D.
Associate Professor
Department of Electrical Engineering Technology
Purdue University

Abstract

Fuzzy cognitive maps were developed as a way to evaluate alternate entry points in complex problem sets where there were many hidden interactions between attributes. FCM's are fuzzy digraphs that map causal linkages between concept nodes. To develop the techniques, the Jasper problem set was used. Participants viewed a video of a search and rescue mission, with relevant information spread throughout the tape. After viewing the tape, the participants, with the help of a facilitator, constructed a fuzzy cognitive map of their reasoning about potential solutions to the problem. Various techniques were used to construct the map, and to evaluate the edge strengths. With a completed map, information could be inferred in one of two ways. For a scenario, its initial conditions could be applied to the map, and a final state for the system determined. In a second way, the edge strengths could be used to define direct and indirect causal linkages.

Examing Alternate Entry Points in a Problem Using Fuzzy Cognitive Maps

Karl Perusich, Ph.D.

Problem Solving

In most real world applications, the potential solution to a problem will require an aggregation of concepts and techniques from many diverse disciplines and knowledge bases. Designing an automobile requires expertise in mechanics, materials, metallurgy, aerodynamics, to name a few. Many technically feasible solutions may exist to the same problem. Additionally, many real world problems require the incorporation of expertise in technically "soft " areas where subjective values by decision makers and final users are primary components in evaluating alternatives. Incorporating the effects of these value judgments with the effects of changing physical parameters usually is not seamless and can be awkward for a designer.

When an attempting to solve a problem, the designer or design team usually exhibits one of two behaviors. Ideally, the practitioners will examine all possible design solutions that meet a given goal within the constraints defined for the problem, and select the one that is "optimal". Optimality is defined by selecting the design that maximizes some cost function that can incorporate objective measures like cost to manufacture or volume of pollutants produced, and subjective measures like consumer preferences or experience levels of users.

Such optimizing behavior is limited in its applicability. Identifying appropriate numerical metrics so that an "apples and oranges" comparison can be made is not easy and not always feasible. Accomplishing a comprehensive search and evaluation of design alternatives for a particular problem can require an unacceptable expenditure of resources

or need an unacceptable period of time to finish. This can be especially true when the design requires a multidisciplinary approach, with tradeoffs in one technical domain affecting the choices of a designer in another technical domain. Very rarely can individuals or resources be found that can effectively identify and evaluate tradeoffs that cross technical boundaries.

A more realizable behavior is one of satificing rather than optimizing. In this case designers try to identify an acceptable solution rather than searching for an optimal solution. A few alternatives are evaluated rather than all alternatives. In very complicated problems requiring many different technical viewpoints, finding a solution may be the only goal. For this type of problem solving boundaries on required resources and design time-frames are more manageable. Satificing behavior still will require "apples and oranges" comparisons to determine if a proposed solution is acceptable when subjective preferences must be incorporated in the evaluation.

When a designer or design team exhibits satificing behavior a search in some sense of possible solutions is still conducted. The designer will use available knowledge, objective and subjective, and propose a solution to the problem. The design will be completed and evaluated for its acceptability. If it meets all constraints and goals, then the process is complete and the design becomes "the" solution to the problem. If it fails to meet the goals and constraints of the problem the search is restarted. The completed design may be re-evaluated to make incremental changes in it to make it acceptable, or substantial parts or all of it may be discarded and alternate methods tried.

The starting point of the process, then, is key to the acceptable solution that is finally arrived at. What particular technology or methodology that is chosen is a function of a variety of attributes and constraints: the technical background of the designers, previous

successful designs, the availability of equipment and resources, to name a few. Different design teams with different backgrounds and different resources can and probably will produce different acceptable solutions to a problem.

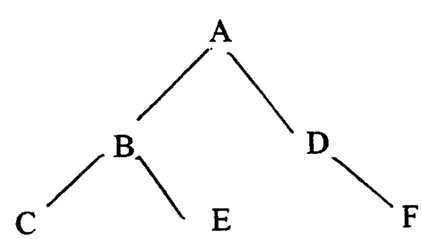
The effect of the starting point on the identification of an acceptable solution or design can be especially acute when the problem is complex, requiring it to be subdivided into smaller problems generally within specific technical domains (electrical, mechanical, computer, etc.). Within each of these problem subdomains, acceptable solutions will be identified that are a strong function of the starting approaches chosen. When these individual solutions are combined the complete solution may be unacceptable, even though each individual solution was acceptable within the subdomain within which it was constructed. An electronic design team may produce a product utilizing a microcontroller. Such a solution may be found to be unacceptable to the manufacturing team involved with the problem because of incompatible equipment, their preferred design would use a PLD, or it may be unacceptable to mechanical engineers assigned to the project because the design produces sufficient heat to need fan cooling unacceptably increasing the size of the unit. In each case acceptable solutions were determined from an initial starting point for the subproblems being examined that were conditioned by the background, capabilities, and resources of the team involved.

Typically, before a design path is chosen for identifying an acceptable solution to a problem, a limited (bounded) search of potential designs is still conducted, but of such limited scope that it can not be considered as optimizing behavior. This bounded search is constrained by the environment in which it is conducted and the way in which it is conducted. The environment constrains the information that the designer can access and use. Methodologies and solutions examined are conditioned by a variety of factors: previous designs-it worked before it should work now, the background of the designers-

EE's trained in analog electronics are not likely to look at microprocessor technology, the available resources-if an optical lab is not in place then the designers are not likely to propose an optical design, the time and money available-with more money and more time a more exhaustive search can be accomplished.

Equally as important as the environment in which the evaluation is done is the manner in which it is done. Very often alternate designs are evaluated in a tree search fashion. A proposed design is evaluated in a sequential fashion against a series of goals and constraints to judge whether it might meet the necessary criteria. A proposed design must meet goal A. If it does, then might it satisfy constraint B? Given that the proposed design meets goal A and might satisfy constraint B, can it meet goal C?, and so on. Several designs may be qualitatively evaluated in such a fashion, with either the first one that meets the criteria chosen, or one of several chosen that meet the criteria using some conflict resolution rule.

Figure 1: Tree Search



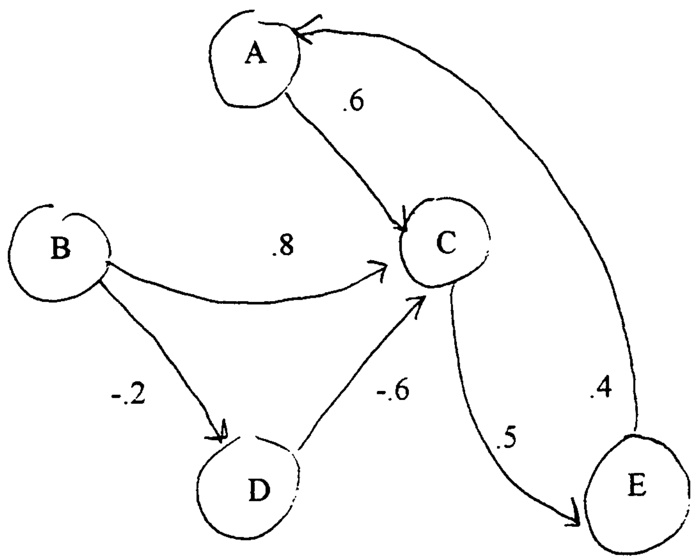
Such an approach can have two serious deficiencies, and, keeping in mind that the initial starting point chosen is critical to the final result, can severely limit the scope of the design. First, a tree-type search ignores potential interactions among goals, constraints, and concepts that can lead to unexpected results. Secondly, such a search can not easily incorporate subjective differentiations that the evaluator may be able to make about the

relative strengths of relationships between various aspects of the solutions and the problem. Since the initial starting point in the development of the solution to a problem is so critical, a better method is needed for qualitatively evaluating the appropriateness of several potential designs.

Fuzzy Cognitive Maps

Fuzzy cognitive maps (FCM) can be used as an evaluation tool to subjectively compare alternate design choices to determine one or several with the best potential to satisfy the goals and constraints of the problem, in essence, to determine the entry point to the solution process. FCM's are digraphs with feedback that relate cause and effect relationships. (In fact, they are sometimes called fuzzy causal cognitive maps.) Each node represents an effect with an edge connecting two nodes indicating a causal relationship. Inference from the map is done in one of two ways. In the first, an initial state (initial set of effects) is used with the map to determine the resulting final state (set of effects). Such inference is akin to receiving answers to what if? questions. In a second technique information is assessed about the hidden relationships that may be present through feedback connections. Even though there may not be a direct cause and effect relationship between nodes A and B (i.e., there would be no direct edge connection between the two), A may indirectly effect B through some alternate path through the map.

Figure 2: Fuzzy Cognitive Map



FCM's have important characteristics that give them great flexibility as a tool for determining the starting point of a design process. They utilize subjective comparisons of states rather than numerical measures of quantities. This allows a variety of dissimilar attributes to be compared. Apples can be compared to oranges. The evaluation can use not only technical attributes like production costs or signal strength, but it can also incorporate social, cultural, political, and economic phenomena that may ultimately alter the acceptability of a solution. Even within purely technical domains, one can evaluate the effects from several technical fields without the need to produce a complete design to make the judgment.

Another important attribute of a FCM is that relative comparisons can be made about cause and effect relationships in the problem. The development of the map relies on the subjective evaluations of experts. As part of eliciting the information such an expert might describe one relationship as very strong, another strong, a third somewhat weak, and so

on. These varying degrees of strength of the relationships are incorporated in the structure of the map by using fuzzy values for the edge strengths connecting nodes rather than crisp ones. Some relationships, then, become more important than others.

A final characteristic is that fuzzy cognitive maps can be combined without each map having to cover the same domain. Just as with the ultimate design process itself, constructing a comprehensive map can be accomplished by subdividing the problem into parts and constructing maps for each part. Separate experts with knowledge specific to a part can be used to construct each submap. The maps can then be joined through common states with the comprehensive map representing a multidisciplinary qualitative model of the problem at hand. By incorporating different viewpoints and different expertise, hidden interactions may result that would not be apparent from the perspective of any single viewpoint. Although fuzzy cognitive maps would add a level of complexity in the problem solving process, they could be of great value in organizing the development of an action plan.

Constructing Fuzzy Cognitive Maps

Before the techniques can be used three important parts must be developed. In the first, an appropriate method must be developed for communicating the technique to the user. At least two approaches are possible. In one the map is preconstructed by the facilitator based on his/her knowledge of the problem. The user is presented with a completed map with predefined nodes and connections. The user then must determine the numerical values for the edge connections to indicate their fuzzy strength. Although this method can help minimize the time required by both the users and the facilitator in constructing the map and may be appropriate when the potential number of nodes gets large or the potential number of experts that must be examined also gets large, it can bias the results

towards the prejudices of the facilitator that constructs the map. Such a map will fail to truly capture the expertise of the expert, which can have serious consequences for the results.

An alternate way and the one preferred is to develop the map with the user (or users). After viewing the tape, a brain storming session is conducted in which the users propose solutions to the problem. As each scenario (solution) is constructed the facilitator should get the participants to explore as many relevant facts and ideas that they can that may impact the particular solution. These scenarios should be discussed in a qualitative way with little reference to any numerical facts that may be available. For each scenario the facilitator would identify cause and effect relationships that have been expounded, and construct a fuzzy cognitive map for the solution. After the map is constructed, the user should be prompted to identify qualifiers for the strengths of the edge connections (causes) using linguistic terms only: very, much, a little, somewhat, a lot, etc., if they have not already done so.

All possible solutions that the users can identify should be examined and FCM's for each constructed. The individual fuzzy cognitive maps can then be combined to form a final map for the solution to the problem being examined. Common nodes across maps become a single node in the final map. It is these common nodes that will provide feedback and the hidden patterns within the structure of the problem. Only individual maps were constructed so there was no need to develop strategies for combining maps from several different experts.

Inferring Information from the Map

Once a fuzzy map has been constructed the knowledge elicited from the expert is captured in the web of connections present and their strengths. The completed map represents a qualitative model of the processes being examined. It can not be used to predict numerical outputs for the system being examined, but instead can be used to assess static changes in states of the system in response to a particular initial state. Given an initial input, the map predicts a final state for the system.

Inferring information about the system being modeled from a fuzzy cognitive map can be done in two ways. In the first, initial state vectors are applied to a map to determine the state that results. Various initial state combinations are tried to find those that result in desirable final state outputs. Solutions to the problem being examined that will not satisfy the goal are eliminated, preventing the expenditure of resources on what would probably turn out to be a dead-end. The remaining solutions can be further paired to find one that has the greatest potential to meet the goals and requirements of the problem. This solution becomes the entry point for the problem solving process.

The expert's knowledge is captured in the map in the connections that are identified. In a second technique for inferring information about the problem from the map, the connection matrix is examined to determine how one node, either directly or indirectly, affects another node. When the causal relationships are given as fractional values, matrix techniques can be used to assign distance between nodes that also incorporate measures of relative strength. With causal strengths represented as fuzzy numbers, as proposed here, these matrix techniques must be modified. As with inferring final states, the matrix math can be modified to act on fuzzy numbers rather than crisp values to produce distributions for the measures of distance and relative strength.

Project Description

Fuzzy cognitive maps were developed as a tool for evaluating alternate entry points in ill-defined problem sets.

Three tasks were accomplished.

1) The methodology of utilizing fuzzy cognitive maps for evaluating entry points in the problem solving process was developed.

2) Although FCM's are grounded on a solid mathematical foundation, in most examples fuzziness is incorporated in the map through edge strengths defined on the interval [-1,1], rather than the values -1,0, and 1. An extension of this was to define the edge strengths by fuzzy numbers rather than fractional values. Such an extension increased the computational complexity of inferring information from the map, but it also increased its flexibility and provided new ways in which information could be elicited.

3) The developed tool was evaluated in a problem solving environment to determine its applicability and utility.

To develop and evaluate the proposed techniques, the Jasper problem set was used. In this problem set, the goal was to find a means of transporting an injured bald eagle from a remote mountain location to a veterinarian some distance away for treatment. The problem and necessary information for its solution are presented to subjects in a short video tape. A variety of transportation modes are available (aircraft, car, foot), as are

several different travel routes. Solutions become combinations of transportation methods and travel routes. Although all information necessary to find an optimal solution is provided, it is spread throughout the video. The challenge is to find one or several routes that get the injured animal to the veterinarian in the quickest time.

The problem solver must determine a proposed solution and then search the video for needed information. For this ill-defined problem there is an optimal solution that gets the injured eagle to the veterinarian in the shortest time. There are a variety of possible travel paths in the problem that might satisfy the goal. Finding the optimal one becomes a search of these different travel paths. As different paths are examined and eliminated, a base of knowledge builds up about the problem requiring less new information each time to evaluate a choice to determine its suitability. Searching for information to assess the suitability of alternate choices represents the key cost to the problem solver. The longer that it takes to find the information to evaluate a proposed travel route, the more it costs to find a solution.

In this scenario the initial route chosen determines how quickly the optimal solution is arrived at. The problem solver will use some qualitative method to assess a starting point for their evaluation of solutions and could be as simple as a "gut reaction". Ideally, the optimal route would be the one examined, eliminating the search of extraneous information, and the consequent waste of time. Fuzzy cognitive maps were used as a way to eliminate unfeasible solutions and assess, qualitatively, the optimality of the remaining solutions.

A methodology was developed for using fuzzy cognitive maps as a tool for qualitatively evaluating the alternate travel arrangements to choose the entry point for finding the solution. Participation in the experiment involved three steps.

First, the subject viewed the video tape. The individual was not allowed to take notes, so all information they used in examining the problem was from memory.

Second, a facilitator constructed a fuzzy cogntive map with the participant about their analysis of the information they had seen. In essence the subject and facilitator were engaged in a brain storming session. In some cases, the participant volunteered a great deal of information about solutions to the problem. In other cases, prompting was required by the facilitator was required. Although the problem and its solution were straight forward, ie. didn't require any specialized engineering knowledge, eliciting the necessary information to construct the fuzzy cognitive map was not always straightforward, especially if the individual had no previous experience with the technique. From this second step the essential states of the problem and their linakages were determined.

To complete the map, edge strengths were determined in the third step. This turned out to be the most difficult step. Several techniques were tried, none being entirely satisfactory. In one, the subjects were asked to assess every edge strength using a rank-ordered set of lingusitic modifiers (adverbs like some, much, a little, etc.) The chief difficulty with this method was the tecndency by the participant to become unfocused. Rather than concentrating on the edge strength being evaluated, they kept evaluating it against other nodes in the map. This tended to produce a map with nearly uniform values for the edge stengths.

A second method was used to assess edge strengths that used modfiied relative comparisons. Transitivty in preferences in a fuzzy cognitve map in general can be assumed. To overcome this handicap, the individual was asked to the relative strengths of all edges affecting the node or all edges eminating from the node. A reference node was

chosen. Relative comparisons were made in a forward chaining/backward chaining (FC/BC) path. These forward chaining/backward chaining search paths would produce a number of distinct chains within the topology of the map. To complete the evaluation, one edge was selected from each chain, typically one that was a minimum or a maximum, and these edges were compared realtively. The edge strengths were then scaled and normalized to give a numerical value on the interval [-1,1]. This method still was cumbersome for some individuals to use, but it did yield a better spread in numerical values. In some maps, it was impossible to choose a FC/BC path that didn't yield some inconsistency in assessed strength. Typically, these could be resolved using follow questions of the participant.

Once constructed the fuzzy cognitive map could be used in two ways to evaluate the entry point for solving the Jasper problem. In the first, a variety of different scenarios could be constructed and applied as initial conditions to the map. From this, a final qualitative state would be inferred from the map. The one with the best qualititive behavior would be entry point.

In a second technique, the edge stregths could be used to identify and value various casual paths in the map, both direct and indirect. Especially of interest would be the indirect paths. Although there may not be a direct casual link between two concepts, there may be one through a mulit-node path. A may not directly cause B, but A might cause D, which causes E, whcih causes B. It is this ability to identify hidden interactions that gives FCM potential value in evaluating complex problems.

The first technique will be the one of primary interest. After the participants have constructed their map it will be used to determine which of their proposed solutions represent the best entry point for the problem solving process. The second technique will

be used in those cases where unexpected behavior results to help understand the processes identified in the map that produced it. With this starting point, the participants will gather the information necessary from the videotape to assess whether the injured bird will in fact get to the veterinarian in time to be saved.

Summary

The techniques of using fuzzy cognitive maps to determine the entry point for solving a problem were developed. In traditional FCM's, fuzziness is modeled using fractional values on the interval [-1,1] to represent the strengths of cause and effect relationships. In this project the concept of fuzziness was extended. Edge strengths were represented by fuzzy numbers (distributions) rather than fractional values. The matrix techniques used for inference were extended to incorporate fuzzy numbers. The thresholding operation that maps nodal values to the state values of -1,0, and 1 in the inference process was replaced by matching the nodal distributions to template distributions for these state values. By utilizing fuzzy numbers, fuzzy hedges techniques could be used in assessing the strengths of the cause and effect relationships. Only one fuzzy membership function would be need to be elicited, with the rest being determined using established mathematical operators for the fuzzy linguistic modifiers used to describe the cause and effect relationships.

With the mathematical techniques established for using fuzzy numbers, the process of using fuzzy cognitive maps for determining the entry point for solving a problem was examined using the Jasper problem set. Participants viewed a videotape describing an ill-defined problem. From an initial viewing, a facilitator elicited information from them and constructed a fuzzy map of their assessment of the problem. From this map potential solutions were tested to identify the one or ones that satisfied the goals of the problem. This solution was then used as the starting point from which the participants proceeded.

A STUDY OF THE ABILITY OF TUNICATES

TO BE USED AS GLOBAL BIOINDICATORS

Dr. Judy Ratliff
Assistant Professor
Department of Chemistry

Murray State University
P. O. Box 9
Murray, KY 42071-0009

Final Report for:
Summer Faculty Research Program
Armstrong Laboratories, Tyndall Air Force Base

Sponsored by:
Air Force Office of Scientific Research
Bolling AFB, Washington DC

and

Armstrong Laboratory

September, 1996

A STUDY OF THE ABILITY OF TUNICATES TO BE USED AS GLOBAL BIOINDICATORS

Judy Ratliff
Assistant Professor
Department of Chemistry
Murray State University

Abstract

The metal content of two solitary tunicates; *Mogula occidentalis* and *Styela plicata*; and two colonial tunicates; *Distaplia bermudensis*, and an as yet unidentified *Didemnum, spp.* collected from the same water system were analyzed for seven metals using Flame Atomic Absorption Spectrometry. Each was found to contain different amounts of Sr, Cd, Zn, Ni, Cr, Cu, and Fe.

The reaction of Styela plicata to elevated levels of chromium was also examined. The chromium studies included the following: 1.) Exposure of the tunicates to different contamination levels to evaluate the toxicity of this metal for this ascidian. 2.) Determination of whether the ascidians accumulated the metal as part of their food source or some other means. 3.) Determination of the extent of metal accumulation in the visceral mass compared to the tunic. 4.) Examination of the ability of the tunicates to depurate the metal from their tunic/visceral mass. 5.) Examination of the palatability of contaminated tunicates to predators and the extent to which the metal contaminant can be passed on to predators.

A STUDY OF THE ABILITY OF TUNICATES TO BE USED AS GLOBAL BIOINDICATORS

Judy Ratliff

Introduction

Many researchers have examined ascidians from different global locations and have reported elevated levels of various metals {1-9}. Most of the reports record the elevated levels of vanadium since the first paper appeared with the amazing discovery of the ability of a living creature to accumulate a toxic substance to levels lethal to most any other organism {10}. This ability of tunicates to accumulate metals has not been thoroughly examined and the potential of these creatures to provide offshore monitoring of metal concentrations has not been fully exploited. Tunicates are found at all depths, in all oceans, marginal seas, and estuaries where salinities exceed approximately 25 ppt. Some are cosmopolitan and can live within a broad climatic range. Many authors have reported varying metal concentrations for ascidians they have examined. Concentrations appear to be different from one area to another and from species to species.

Each tunicate filters large volumes of water and accumulates/tolerates a wide variety of metals, they are sedentary, abundant, easy to collect, hardy enough to survive and reproduce under laboratory conditions, and large enough to provide adequate sample sizes for analysis. The concentration of trace elements within the ocean varies considerably with location. Furthermore; the ability of ascidians to accumulate levels of metals beyond what is toxic to current bioindicator organisms increases their value as sentinel organisms.

Metals of primary interest to our study were those which appeared on a list of 65 toxic pollutant groups {11}. These elements which appeared singly and combined are shown in Table 1.

Methodology

The majority of the specimens used in this study were collected from ropes that had been tied to a pier in the St. Andrew Bay where they had settled (*Styela plicata* and *Distaplia bermudensis*). The remainder were collected off the shore of Crooked Island (*Mogula occidentalis* and the *Didemnum spp.*) The studies were set up as described below:

Study 1; The determination of the concentration of selected metals found in the chosen tunicates, raw sea water, and controls taken from the study area. Samples were digested as explained below and analyzed with a Varian SpectrAA (Varian Techtronics, Ltd.; Springvale, Australia)

Study 2; Response of *Styela plicata* to varying concentrations of chromium. This tunicate was chosen to begin with because of previous experience of the researchers in handling this species in aquaria kept in the laboratory. The aquaria were cleaned with a commercial dishwashing liquid, scrubbed with a synthetic sponge, rinsed ten times with tap water after all traces of soap had been washed away then taken to the 9700 area pier. At the pier the aquaria were rinsed ten times with raw sea water then filled with the sea water and left

List Priority	Element	Concentration in Sea Water (ppm) {12}
5	Antimony	2.4×10^{-4}
6	Arsenic	3.7×10^{-3}
11	Cadmium	1×10^{-4}
21	Chromium	3×10^{-4}
22	Copper	1×10^{-4}
44	Lead	5×10^{-7}
45	Mercury	3×10^{-5}
47	Nickel	1.7×10^{-3}
56	Selenium	2×10^{-4}
57	Silver	2×10^{-6}
60	Thallium	1×10^{-5}
65	Zinc	5×10^{-4}

Table 1: Metals considered singly and combined as toxic pollutants.

soaking in it for greater than eight hours. This was done to allow any contaminates that might still remain on or absorbed by the polyethylene aquaria to leach out into the sea water. After approximately eight hours the aquaria were emptied then rinsed ten more times with fresh raw sea water. They were then taken back to the lab to await the addition of the tunicates. Returning to the pier Styela plicata were collected from the ropes, scrubbed as clean as possible with a toothbrush then rinsed and taken to the lab along with three carboys of raw sea water. The following aquaria were then set up; Half contained raw sea water and half filtered sea water - this was done to try to determine if any metal that might be accumulated was dependant on the presence of the phytoplankton food source or not. The filtered sea water was obtained by passing raw sea water through 0.45 micrometer cellulose nitrate filters.

Raw Sea Water	Filtered Sea Water
Controls	Controls
1.0 ppm Cr	1.0 ppm Cr
10.0 ppm Cr	10.0 ppm Cr
100.0 ppm Cr	100.0 ppm Cr

Water was changed daily and the following samples were collected; 1.) Raw sea water, 2.) Filtered sea water; 3.) Unfiltered aquarium water (prior to and after water changes); 4.) Filtered aquarium water (prior to and after water changes); and 5.) Feces from the aquaria. When tunicates were analyzed, whole animals were analyzed as well as separating selected samples into tunic and visceral mass.

Study 3; 24-Hour Chromium Time Studies were carried out to analyze the rate of chromium accumulation and determine the time at which the maximum amount of chromium had been accumulated by the tunicate. The aquaria were prepared as before but this study placed two tunicates in each aquaria. None of the water was filtered since no significant difference in chromium accumulation levels had been noted in the previous study. All aquaria were spiked with 10.0 ppm Cr. *Styela plicata* were exposed to 0, 1, 2, 4, 6, 12, and 24 hours of the spiked solutions then harvested, dissected, rinsed, and analyzed.

Study 4; Environmental Implication of Chromium Exposure. The aquaria were prepared as before then the tunicates were collected and placed into the aquaria as follows; one third were used as controls, one third were exposed to 0.1 ppm Cr, and one third was exposed to 1.0 ppm Cr. Water was not changed for one week in any of the aquaria. After this time; One set of 0, 0.1, and 1.0 ppm Cr was harvested and analyzed. One set was left untouched, one set was placed in fresh raw sea water and had their water changed daily for the rest of their lives. This last set was to determine if the accumulated chromium could be cleansed from the tunicate or if it became a permanent part of the tunic. The remaining set was placed with a predator (*Melongena corona*). This was done to determine if the chromium

made them less tasty to the predator and if perhaps this contaminant could be passed on along the food chain.

Digestion

Samples were dried at 60-70°C for 48 hours and a dry weight was obtained prior to tearing each into small pieces. Next about 100 mL of deionized (18 MΩ Milli Q®) water and 5 mL of concentrated nitric acid was added to each sample which was then covered with a ribbed watch glass. If the sample was large and not completely covered by the water/acid mixture, more water was added. The ribbed watch glasses direct condensate back into the beaker, protecting the sample from contamination from the outside, and serve as a splatter guard for a bubbling sample. Boiling beads were then added to the mixture to help prevent bumping. Greater amounts of sand found in the tunic led to a greater problem with samples having a tendency to bump.

The samples were then evaporated to about 25 mL on a hot plate, without boiling, liquid levels were kept above the sample so as not to let any part of it dry out. Samples were then cooled and 5 mL more of HNO_3 added. The nitric acid is Fisher A509-212-Trace Metal Grade. Samples were then gently refluxed without allowing them to boil dry at any point in time or part of the beaker. Acid was then continually added and samples refluxed until digestion was complete.

Once digestion was complete samples were evaporated to about 15 mL without allowing the beaker to become dry at any point then 10 mL of concentrated HCl was added. The hydrochloric acid used was Fisher A508-212-Trace Metal Grade. Filters took one hour to digest, fecal matter required one day, and tunicates typically took two days per species to digest.

The beaker walls and watch glasses were then washed down into the sample beaker and the entire sample diluted to 100 mL. Samples were then analyzed using a Varian SpectrAA, Flame Atomic Absorption Spectrometer.

All water samples were analyzed directly, without digestion. Blanks were carried through all analyses to correct for possible contaminants present in the filters, acids, or water used in the digestion. Interference from iron for the chromium was checked separately, first by adding ammonium chloride and then by using the method of standard additions. The method of standard additions will normally detect interference from any source, not just Fe. No interference has been seen.

Filters used were: Metrical Membrane Filters, 0.45 µm and Membrane Filters, 100% Cellulose Nitrate, 0.45 µm, Whatman®, Whatman International Ltd., Maidstone, England.

Filter Handling:

Filtered samples were collected by removing a filter from their box with plastic coated forceps and placing it on the fritted filtration support. The filtration apparatus was then assembled and the sample passed through. The filter paper itself was then folded in half, over onto itself and placed in a small, zip loc, polyethylene bag once again using the plastic coated forceps with the aid of a plastic coated spatula. The water sample that had been filtered through was then poured into a polypropylene or high density polyethylene bottle and labeled for subsequent analysis. The filtration apparatus was rinsed at the end of a

collection cycle with tap water and allowed to dry. Prior to its use it was rinsed with the sample being filtered.

Samples which contained Cr were rinsed with approximately 10.0 mL of filtered sea water prior to storing in their bag. This was done to prevent osmotic pressure from rupturing of any sea organisms and the subsequent loss of said organism through the filter.

The filters were then removed them from their bags with the plastic forceps and placed them in a beaker. The bag was then rinsed 10 times with 18 MΩ water with the rinse poured into the beaker with the filter. The samples were then digested and analyzed as outlined above.

Results/Discussion

Metals Survey

A survey of the metal content of the tunicate samples was required to determine which metals the animals naturally accumulate. This is not an absolute analysis of the metals they *can* accumulate, since if a metal contaminate is not present in their environment they cannot accumulate it. The metals and samples surveyed (that time permitted) were strontium, chromium, iron copper, nickel, cadmium and zinc. The concentration of these metals was determined in the four tunicates examined, *Styela plicata*, *Mogula occidentalis*, *Distaplia bermudensis*, and the *Didemnum spp.* Since shellfish are commonly used as bioindicators of the metals of interest the concentration of the metals in *Atrina spp.*, *Crassostrea virginica*, and *Spisula solidissima* along with a mixed handful of seagrasses, raw sea water, and the fleshy part of *Melongena corona* was also determined. The results of these analyses are summarized in Table 2.

Figure 1 shows clearly that the *Didemnums spp.* are the better accumulators of the majority of the metals examined, though one species is not responsible for the greatest accumulation all the time, they - as a group - accumulate chromium, iron, nickel, and strontium better than any other samples analyzed. Copper was accumulated best by the fleshy part of *Melongena corona* and the mixed sea grasses. Cadmium appears to be best accumulated by *Styela plicata*. Since the concentration of each of these metals was determined using Flame Atomic Absorption Spectrometry (FAAS) it was very time consuming. If a greater number of metals were to be examined and a greater number of tunicates, the best , most cost efficient, and time effective method would be to use Inductively Coupled Plasma Atomic Emission Spectrometry (ICP-AES).

If the enrichment factors for these metals in the tunicates are compared to the enrichment factors of shellfish commonly used as bioindicators their abilities are again highlighted as seen in Table 3. In each case one of the extremely small number of samples surveyed was capable of accumulating more metal per gram dry weight than the shellfish. The shellfish analyzed in this study have lower enrichment factors than those in the literature. The observed depression is most likely because the survey scans were done using animals that were taken from a relatively clean bay and analyzed; little or no contamination was present in the water being filtered by these animals. Typical enrichment factors are calculated using animals which have been exposed to a metal contaminant.

Since the tunicates examined are stationary, attaching themselves to pier pilings, ropes, rocket

Animal	[Sr] (ppm)	[Cr] (ppm)	[Fe] (ppm)	[Cu] (ppm)	[Ni] (ppm)	[Cd] (ppm)	[Zn] (ppm)
Raw Sea Water	7.07	ND	ND	ND	0.15	0.04	0.06
Raw Sea Water	6.99	ND	ND	ND	0.12	0.04	0.05
Atrina rigida	674.86	0.33	368.59	8.29	5.15	2.28	163.96
Atrina serrata	463.71	ND	210.78	6.74	7.52	1.69	159.43
Atrina siminuda	574.85	ND	349.21	6.55	10.23	2.33	199.65
Crassostera virginica	1184.86	4.67	169.7	0.73	8.12	1.57	1010.9
Didemnum sp. #1	1460.54	193.52	1687.01	3.39	11.8	ND	318.51
Didemnum sp. #2	4980	ND	907.5	ND	222.5	ND	112.5
Didemnum sp. #3	7525.29	14.7	365.46	2.67	8.11	1.23	39.8
Distaplia bermudensis	47.75	1.41	0.69	2.82	3.01	0.68	23.95
Melongena corona	28.6	3.89	1.25	50.8	2.98	0.7	30.55
Mogula occidentalis	33.21	8.76	649.13	0.83	2.35	0.24	6.92
Mixed Sea Grasses	38.35	8.94	157.25	13.97	4.9	0.38	8.9
Spisula solidissema	1711.02	2.4	98.81	0.91	10.3	1.84	39.15
Styela plicata	ND	ND	326.44	ND	20.24	4.24	19.6

Table 2: Summary of metals analysis. All numbers given are averages if multiple samples were analyzed.

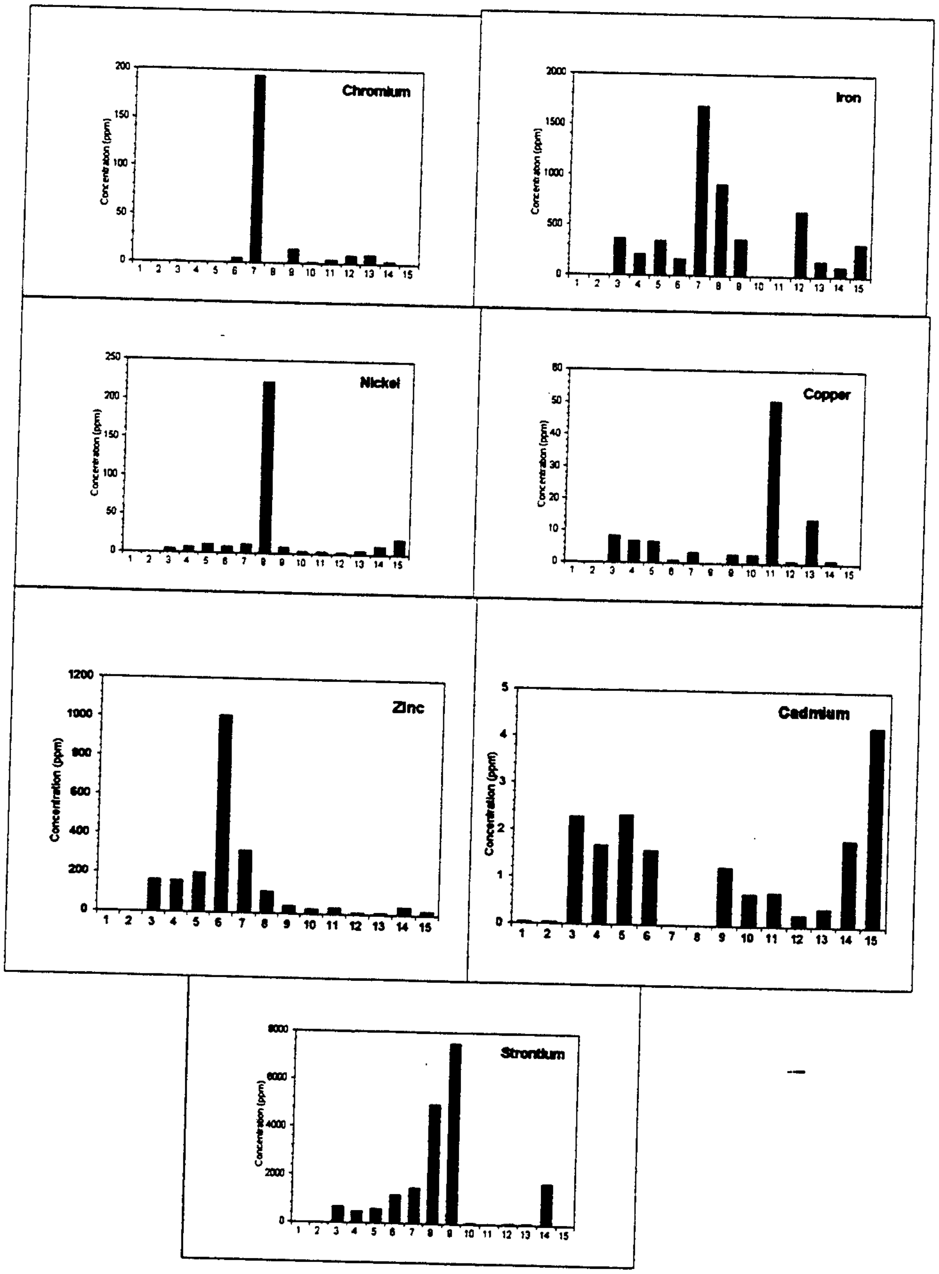


Figure 1: Metals accumulation data for species examined. 1 & 2 are Raw Sea Water, 3-5 are the *Atrinas spp.*, 6 is *Crassostera virginica*, 7-9 is the *Didemnums spp.*, 10 is *Melongena corona*, 11 is *Mogula occidentalis*, 12 is the mixed seagrasses, 13 is *Spisula solidissema*, and 14 is *Styela plicata*.

Element	Concentration in Sea Water (ppm)*	Scallop*	Oyster*	Mussel*	*D. bermudensis*	*S. plicata*	*M. occidentalis*	*Didemnum spp.*
Sr	8.0E+00				6.0E+00	ND	4.2E+00	5.8E+02
Cr	3.0E-04	2.0E+05	6.0E+04	3.2E+05	4.7E+03	ND	2.9E+04	2.3E+05
Fe	2.0E-03	2.9E+05	6.8E+04	2.0E+05	3.4E+02	1.6E+05	3.2E+05	4.9E+05
Cu	1.0E-04	3.0E+03	1.4E+04	3.0E+03	2.8E+04	ND	8.3E+03	2.0E+04
Ni	1.7E-03	1.2E+04	4.0E+03	1.4E+04	1.8E+03	1.2E+04	1.4E+03	4.8E+04
Cd	1.0E-04	2.3E+06	3.2E+05	1.0E+05	6.8E+03	4.2E+04	2.4E+03	4.1E+03
Pb	5.0E-07	5.3E+03	3.3E+03	4.0E+03	1.1E+07	ND	1.9E+06	2.3E+07
Zn	5.0E-04	2.8E+04	1.1E+05	9.1E+03	4.8E+04	3.9E+04	1.4E+04	3.1E+05

Table 3: Enrichment factors of shellfish and tunicates compared with the marine environment.

boosters, grasses, etc. they are better indicators of local concentrations of metals than the mobile shellfish. If a tunicate accumulates a metal, the metal must have been present in the water at the tunicates location at some time. Shellfish are mobile, they can propel themselves along or be easily carried by ocean currents to different locations - therefore, when they are collected it can not be definitively said that the metals accumulated by the shellfish arose in that location.

Exposure Concentration Studies

To further explore the response of tunicates to metal contaminates a series of studies were designed. Since the scans had not been completed when these studies were initiated, chromium was selected as the metal of interest to be examined first. This choice was based on the current attention being received by chromium VI and the fact that when we began out study the only atomic absorption lamp we had was a chromium lamp.

The response of *Styela plicata* to exposure to varying concentrations of chromium was carried out as described in the methodology for Study 2. *Styela plicata* exposed to 100 ppm chromium died within 24 hours, 10 ppm chromium within 48 hours, and 1 ppm lived indefinitely.

Upon analysis of the samples collected in this study it was determined that the greatest concentration of chromium was in the tunic of the ascidian. There was no chromium detected in the sea water, raw or filtered; no chromium detected in any of the controls used in this study - visceral mass or tunic; and no chromium detected in the fecal matter of controls or Cr spiked *Styela plicata*. Chromium concentrations in the aquaria did decrease from day to day, indicating the uptake of the contaminate by the ascidian but since the aquaria were not covered, the rate of evaporation of water from the aquaria was not constant and the exact amount could not be determined. A summary of the accumulation of chromium is presented in Figure 2. The tunics do show a greater amount of chromium present than within the whole animal in all cases except for one. (Since time did not permit repeating the study beyond the duplicate samples run in each analysis it was easy for one data point to skew the results.) There was no significant difference seen between the tunicates exposed to chromium in filtered versus raw sea water, this indicates that the chromium uptake is not dependant upon the phytoplankton food that is filtered by the tunicate.

These tunicates are filter feeders who pass sea water through a mucus membrane having openings of approximately 0.5 micrometers. The filtered sea water used in the aquaria was filtered using a 0.45 micrometer filter - therefore, any phytoplankton caught by the *Styela plicata* as food should have been filtered from the water the tunicate was housed in. Since tunicates maintained in filtered sea water were able to accumulate similar amounts of chromium as those kept in raw sea water the method of accumulation of chromium can not be dependant on the phytoplankton food source.

An overall increase in the concentration of chromium accumulated was noted when the Styela plicata were exposed to different concentrations of chromium as shown in Figure 2. To more closely examine this discovery the following study was designed and carried out.

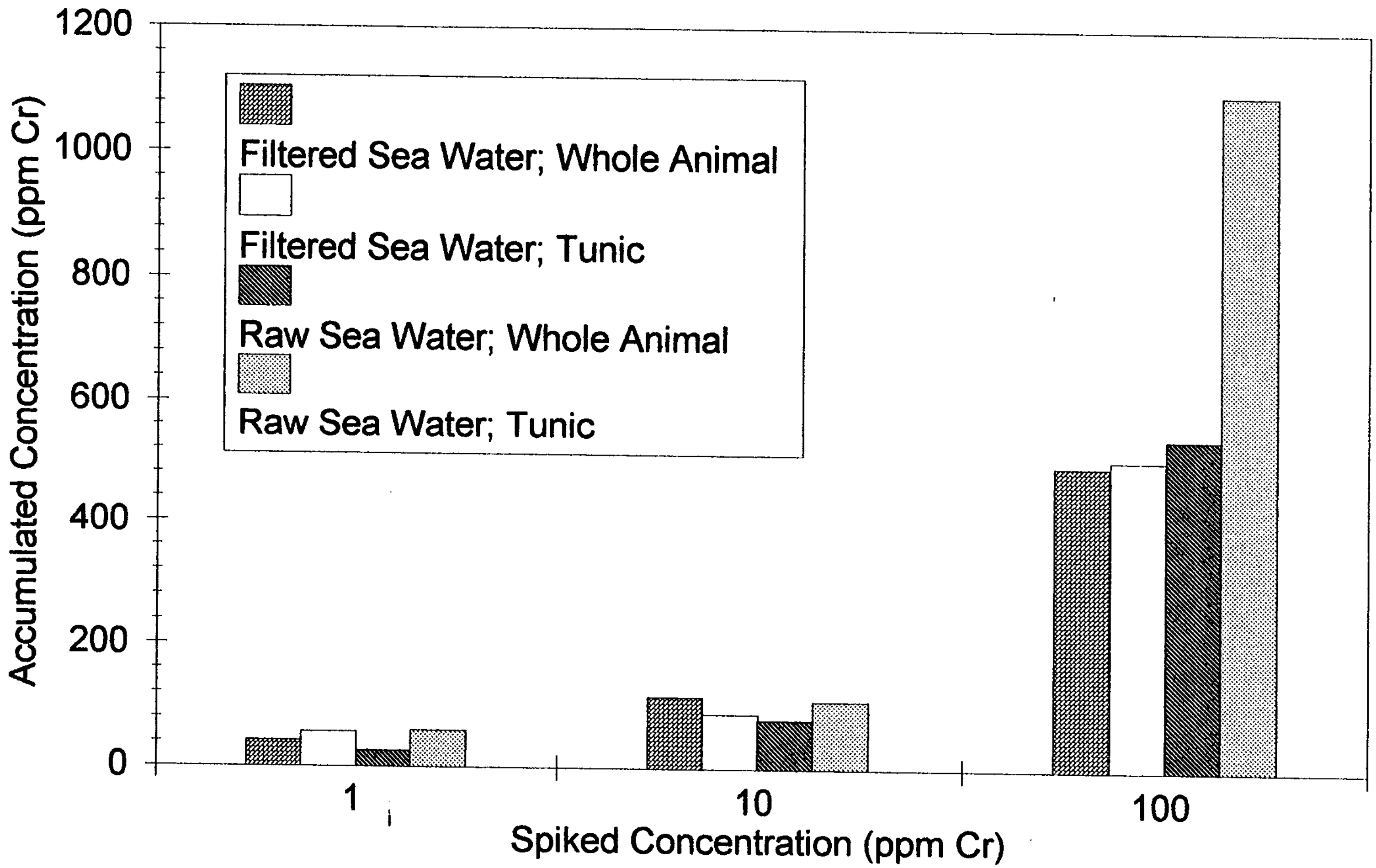


Figure 2: Concentration studies with *Styela plicata* being exposed to 1, 10, and 100 ppm chromium.

Time Studies

To determine the length of time for the tunicates to accumulate a maximum amount of chromium a series of experiments were set up as explained in Methodology, Study 3. The data obtained is summarized in Table 4. The concentration of chromium accumulated increases in both the tunic and visceral mass as the length of exposure increased. However, as seen in Figure 3 the rate of increase was not the same. The concentration of chromium in the tunic appeared to level off after six hours but the concentration of chromium in the visceral mass continued to increase linearly. Since the Styela plicata were known to die after exposure to 10 ppm chromium after 48 hours, death may result once the level of concentration in the visceral mass reaches a toxic level for the tunicates. It could also be hypothesized that the level of concentration of chromium in solution was absorbed too quickly by the visceral mass and the Styela plicata did not have time to be regulate the amount being taken in; perhaps transferring the bulk of the contaminant to the tunic. Tunicates exposed to 1 ppm chromium however, would have more time to transfer absorbed chromium to the tunic and away from vital organs, keeping the Cr levels in the visceral mass at a less than toxic level.

Environmental Implication Studies

All *Styela plicata* samples in this study, including the controls appeared to have accumulated more chromium than the previous studies as seen in Table 5. Leaving the tunicates in the chromium spiked solutions for a week generated some opposite results of those who were given fresh (yet chromium spiked) sea water each day. The visceral mass accumulated more chromium than the tunic for all samples that lived until their predetermined harvest date. Three who died earlier than the harvest date had accumulated more chromium in their tunic than those who remained alive until harvest as had been seen in the earlier study. However, once again the greater accumulation levels were directly related to higher exposure levels.

The *Styela plicata* used in this study were larger ones of approximately equal size. They were selected this way to try to keep as many of the conditions we could control the same. Large animals were chosen because many literature references suggest they are better accumulators of metals than smaller ones. It seems reasonable to assume that since the aquaria the *Styela plicata* were kept in for this study were crowded, not aerated, and none had their water changed for one week that they may not have been in good health. For one week they would have been cycling the same water through their bodies, including their own waste products, and competing for any available air. This may have hindered their ability to handle contaminants, such as the chromium, which they were exposed to.

Depurating Study

In examining the data in Table 5 few trends can be seen when comparing tunicates that were left in the aquaria and those that were placed in fresh sea water every day until their harvest. However, if the concentration of chromium in the tunics of those animals exposed to 0.1 and 1.0 ppm chromium are compared - the concentration of chromium is generally lower in the tunics of the depurating *Styela plicata*.

Kill Time (Zulu)	Exposure Time (Hrs.)	Vol. Disp. (mL)	Dry Wt. Tunic (g)	Dry Wt. Viscera (g)	Conc. Cr Tunic (ppm dry wt.)	Conc. Cr Viscera (ppm dry wt.)
1000	0	80	2.65	1.7	5.56	ND
1100	1	100	4.09	2.29	10.84	ND
1100	1	105	3.72	1.98	22.47	ND
1200	2	100	3.4	2.24	45.38	ND
1200	2	110	4.66	2.29	23.13	6.15
1400	4	80	3.48	2.18	34.86	ND
1400	4	95	3.81	2.5	26.69	5.74
1600	6	85	3.7	1.57	43.81	8.5
1600	6	80	3.45	1.8	39.88	12.19
2200	12	100	3.75	2.13	31.31	11.9
2200	12	90	3.79	2.05	43.51	ND
1000	24	80	3.59	2.3	55.46	0.042
1000	24	85	3.1	1.44	45.74	46.48

Table 4: Concentration of chromium in *Styela plicata* after different lengths of exposure. Two tunicates were housed in each aquarium. The concentration of chromium in all aquaria was 10 ppm.

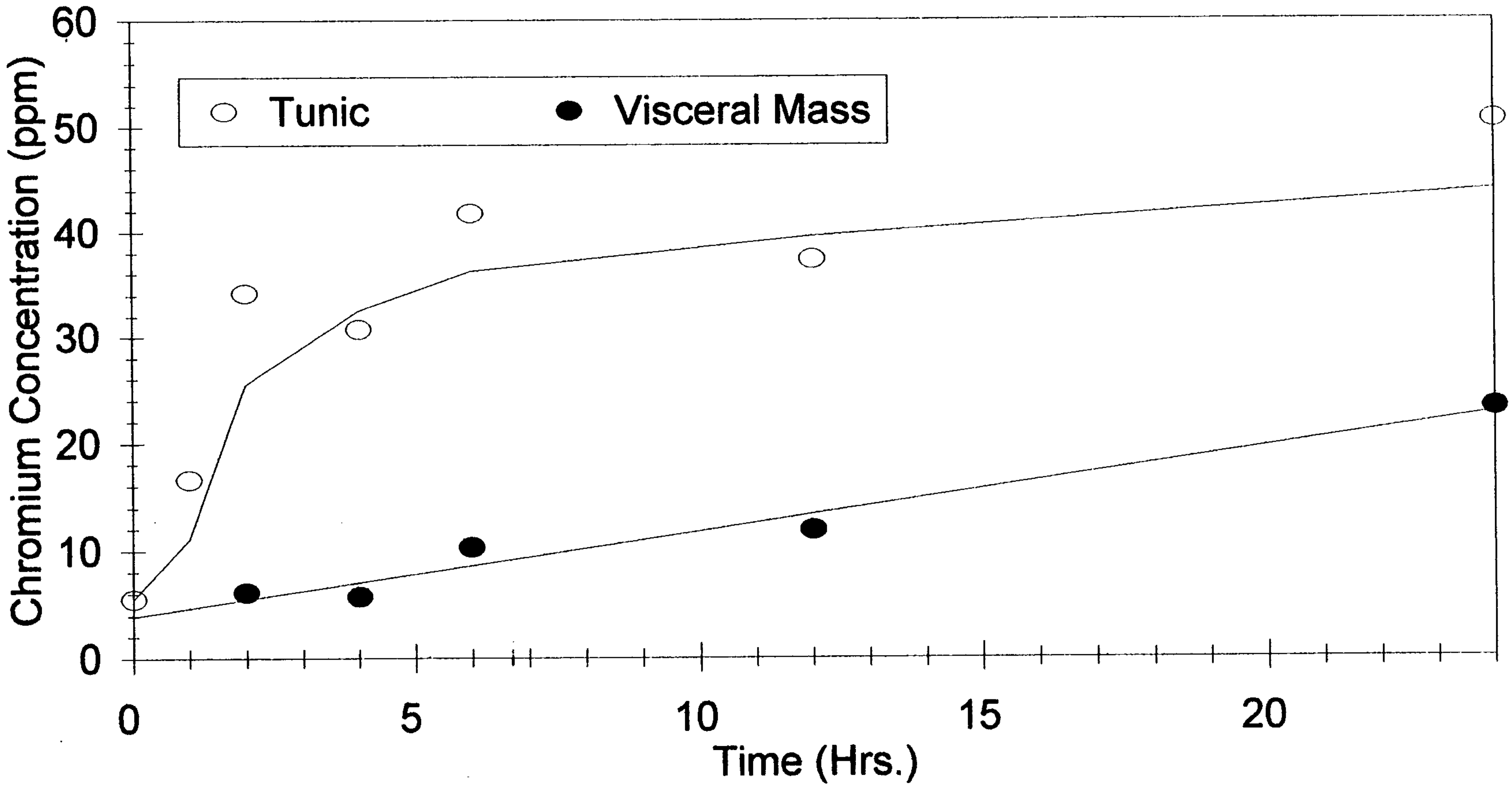


Figure 3: Concentration of chromium in *Styela plicata* after different lengths of exposure. The concentration of chromium in the tunic levels off after approximately six hours but that of the visceral mass increased linearly throughout the study.

[Cr] Spike	[Cr] Tunic	[Cr] Visceral Mass	Use
0	19.73	34.41	Depurating study
0	14.93	16.63	Depurating study
0	2.13	5.49	
0	4.34	5.07	
0.1	10.5	10.43	Depurating study
0.1	9.05	32.55	Depurating study
0.1	14.66	22.41	
0.1	11.75	15.93	
1	31.81	48.86	
1	53.39	19.87	Died after 1 week
1	52.14	56.86	
1	25.5	46.51	Depurating study
1	36.73	86.02	Depurating study
1	57.39	71.1	
1	73.18	69.15	

Table 5: Styela plicata spiked with the indicated concentrations of chromium were left for approximately two weeks in untouched aquaria . The ones labeled as Depurating had their water changed daily for the last week of their life and replaced with fresh raw sea water (no chromium contamination was added to this solution).

This trend was not seen in the visceral mass of these same specimen; if anything the opposite was seen. No explanation will be attempted at this time.

Appeal to Predators

The *Styela plicata* eaten by the *Melongena corona* were analyzed whole after the attack, once the snails had relinquished their hold. The visceral mass was the only noticeable material being eaten by the predator and varying amounts of this material was left. No difference in the amount or appearance of the remaining visceral mass was noted in the Styela plicata eaten by the Melongena coronas. The average concentration of chromium in the ascidians that were eaten were 5.43 ppm for the controls, 12.46 for the 0.1 ppm spike, and 34.7 ppm for the 1.0 ppm spike. These concentrations represent what would be expected of the concentration of chromium in the tunic. Since the visceral mass left was extremely small in comparison to the tunic, it seems reasonable to assume it contributed very little chromium.

No difference was noted in how long it took the *Melongena* corona to attack the contaminated *Styela plicata*. In fact a tunicate exposed to the 1.0 ppm Cr was attacked first and one that had been a control with no exposure to chromium was left untouched by all of the Melongena it was placed in contact with, including starved ones.

When the Melongena corona that had eaten the chromium contaminated animals were analyzed, no trend in the concentration of chromium in the fleshy part was found, the shell was not analyzed. The feces of the snails, did however, contain greater concentrations of chromium. The average concentration in the snails was around 4 ppm but that in the fecal matter analyzed was 85 ppm. This suggests that the contaminant was not passed on to the predator.

Future Research Suggestions

These studies indicate that ascidians are sentinel organisms that are already in place in all oceans and marginal seas. Their presence in these locations could easily be exploited to monitor pollution levels as well as manufacturing processes being carried out by our neighbors that might effect our waters. To use them most effectively, their response to metals of interest needs to be mapped. This initial study has surveyed an extremely small number of tunicates and only scratched the surface in examining their metal accumulation abilities.

References

1. Ogawa, Nobuaki; Nakamura, Akira; Sato, Hiromitsu; Onoue, Takashi; Yamamoto, Masahiko; Matsuo, Shigeki The Chemical Society of Japan, 1991, 293.

2. Cossa, Daniel Oceanologica Acta,12, 1989, 417.

3. Smith, M. J. Experientia, 45, 1989, 452.

4. Danskin, Gordon P. Can. J. Zool., 56, 1978, 547.

5. Bruening, Reimar C.; Oltz, Eugene M.; Furukawa, Jun; Nakanishi, Koji; Kustin, Kenneth J. Am. Chem. Soc., 107, 1985, 5298.

6. Levine, Estees P. Science, 133, 1961, 1352.

7. Petersen, Kjerulf J.; Rusgard, H.U. Mar. Ecol. Prog. Ser., 88, 1992, 9.

8. Papadoupoulou, Catherine; Kanis, George D. Mar. Pol. Bul., 8, 1977, 229.

9. Roman, Domingo, A.; Molina, Justa; Rivera, Lidia Biol. Bull., 175, 1988, 154.

10. Henze, M. Hoppe-Seyler's Z. Physiol. Chem. 72, 1911, 494.

11. 40 CFR 401.15, National Bureau of Affairs, Inc.

12. The Open University, P.O.Box 48, Walton Hall, Milton Keynes MK7 6AA England. Seawater: Its Composition, Properties and Behavior, Pergamon Press, Inc., Elmsford, New York, 1989, pp, 86.

Computerized neuropsychological assessment of USAF pilots.

Paul D. Retzlaff
Professor
Department of Psychology

University of Northern Colorado
Greeley, CO 80639

Final Report for:
Summer Faculty Research Program
Armstrong Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, DC

and

Armstrong Laboratory

July 1996

Computerized neuropsychological assessment of USAF pilots.

Paul D. Retzlaff
Professor
Department of Psychology
University of Northern Colorado

Abstract

The neuropsychological assessment of US Air Force pilots presents several unique problems given their relatively high cognitive functioning. The US Air Force currently has a baselining procedure wherein student pilot candidates undergo computerized cognitive assessment. The intent of this assessment is to archive pre-morbid data against which to compare potential future post-accident performance. The current work provides the necessary background, clinical methods, and data in order to assess pilots who have suffered cortical insult such as trauma, disease, or toxin exposure. Methods are delineated for those with pre-morbid testing as well as for those pilots without such testing.

Computerized neuropsychological assessment of USAF pilots.

Paul Retzlaff

Aviation is one of the most cognitively demanding occupations. Any decline in cognitive ability is of great concern from a number of perspectives. After initial flight training, a number of cognitive insults may result in an occupationally significant cognitive decline. These insults can include chronic alcohol abuse, brain trauma, cerebrovascular insufficiencies, neurodegenerative diseases, and psychiatric disabilities such as depression. The resultant declines in performance may be temporary or permanent. The complexity of aviation jobs and the unforgiving nature of the working environment demands a conservative approach to an occupational return after even the smallest central nervous system insult. At a minimum, medical and neurological evaluations are completed, but in addition, neuropsychological assessment may be indicated.

The purpose of the present paper is to provide clinical procedures for the evaluation of pilots with cognitive referral questions and to provide the necessary comparative test norms. Procedures are provided for patients who have pre-morbid EFS testing and for those without such testing.

METHOD

Subjects

A sample of 537 Air Force pilot training candidates participated in this study. The sample as a whole had a mean age of 23.5 (sd 4.2) and about 8% were female. Subjects who had been commissioned through Officer Training School, ROTC, and the Air National Guard were all college graduates. Approximately, 42% were Juniors at the United

States Air Force Academy. Student pilot candidates participated in the baseline cognitive testing during EFS either at the Air Force Academy in Colorado Springs, CO, or at Hondo, TX.

Measures

The Multidimensional Aptitude Battery (MAB) (Jackson, 1985) is a broad based test of intellectual ability. It was patterned after the Wechsler Adult Intelligence Scale (WAIS-R; correlation = .91), the most widely used individually administered test of intelligence. While the WAIS-R requires about an hour and a half per subject to administer, the MAB can be given to groups and requires about the same amount of total testing time. Additionally, the WAIS-R requires skillful scoring while the MAB has a multiple choice format. All subtests in the WAIS-R have corresponding paper and pencil subtests in the MAB except immediate digit memory. Verbal components tapped include information, comprehension, arithmetic, similarities, and vocabulary. Performance measures include digit symbol coding, picture completion, spatial, picture arrangement, and object assembly. Scores on each of the subtests are scaled to a mean of 50 and a standard deviation of 10. Verbal and performance sub-scores are available as is a full scale intelligence score, each scaled to a mean of 100 and a standard deviation of 15. Reliabilities for the summary scores range from .94 to .98.

Current testing in the USAF Enhanced Flight Screening program (King and Flynn, 1995), other US Air Force research programs (Flynn, Sipes, Grosenbach, and Ellsworth, 1994; Retzlaff and Gibertini, 1988), NASA's astronaut selection procedure, and a number of civilian airline

screening procedures include the MAB.

The version of the MAB used in the current study was primarily the Armstrong Laboratory's computerized version (Retzlaff, King, and Callister, 1995). Here verbal questions are presented as text on a computer screen and subjects are asked to respond to the computer with an a, b, c, d, or e keyboard entry. The performance items were scanned into computer graphic files and are presented in a window on the monitor. This computerization was done and is used with the consent of the test author with explicit copyright permission. It is important to note that the 1990 norms for the MAB were used for this study. These norms are used in the computer scoring software from the publisher. Earlier work with the test or other current paper-and-pencil type administrations use the original 1985 norms. Hence, direct comparison with data such as Retzlaff and Gibertini's (1988) may be difficult.

The CogScreen-Aeromedical Edition (Kay, 1995) is a test of cognitive ability intended for use in the assessment of pilots. While the MAB is a test of relatively complex, higher order intellectual processes, the CogScreen tasks are generally more fundamental processes such as reaction time. It is not a test of aviation knowledge but considered to include abilities necessary in the performance of aviation duties (Kay and Horst, 1988; 1989). There are 11 tasks which result in 65 scores. The tasks include Backward Digit Span (BDS), Math (MATH), Visual Sequence Comparison (VSC), Symbol Digit Coding (SDC), Matching-to-Sample (MTS), Manikin (MAN), Divided Attention (DAT), Auditory Sequence Comparison (ASC), Pathfinder (PF), Shifting Attention (SAT), and Dual Task (DTT). Each of the tasks is

usually scored in a number of ways. Typical scorings include task speed, accuracy, and throughput. Throughput is a function of speed and accuracy, basically the number of correct responses per minute. It is indicative of the amount of work accomplished. A number of tasks also include process completion measures which quantify task specific behavior such as control of the computer screen elements. The manual and other research refers to the CogScreen scores by a relatively cryptic variable naming process.

The CogScreen is relatively new and represents an attempt by its authors to produce an assessment device which met a number of FAA requirements. It is currently used in the EFS program by the USAF, by the US Navy, and by a number of commercial airlines. It is published and available from one of the major psychological test publishers.

The CogScreen was used as provided by the test publisher. Software administers the test, times the tasks, scores the tests, and archives the data in report form.

Clinical methods

There are three major manners in which to use the available data (Retzlaff and Gibertini, 1994). The first is the intended purpose of EFS. This procedure compares the archived data (pre-morbid) to later testing (post-morbid), presumably after some sort of cognitive insult.

The other two procedures acknowledge the fact that not all pilots will have archived pre-morbid data. This may be the case because either they became pilots before the program began or they become pilots after the program was terminated (if indeed the program is

terminated). These two procedures use data developed from those taking the EFS testing. As such, the second procedure looks at the relative ability level of the new patient given the known ability levels for the tested group. The third and final method uses a number of the tests for a new subject as control conditions for other tests taken at the same time.

Change in Performance Method The first way is a pre-test, post-test method. It is the most reliable but requires prior, pre-morbid testing data against which to compare later testing. In the general clinical case, a patient may have prior intelligence and neuropsychological testing, been exposed to some cortical insult, and then re-tested. An example might be a patient in the Veteran's Administration system. It would be common for a patient to have a prior intelligence test such as a WAIS-R somewhere in the system, have some sort of cortical insult such as a stroke or head injury, and then be re-tested on the same intelligence test. Here the results of the first testing can be used as a reference for the second testing. A significant decrement across testings would establish the existence of a dementia and gauge the general severity of it.

The degree to which test scores may vary from one testing to the next can be established statistically. "Normal" or chance degrees of differences can be established through studying the stability of normal subjects across two testing periods. The first testing is correlated with the second to establish a stability (reliability) coefficient. This coefficient can be used to determine a confidence band around a score. Performance beyond this confidence band would suggest performance decrements beyond what might be expected by

chance.

For aviators who have participated in the EFS program, pre-morbid data is available and can be retrieved from Armstrong Laboratory. Knowing the aviator's initial performance, the stability coefficient of the test, and the variability of the test for aviators, confidence bands can be established for an individual aviator. Performance below what can be expected statistically on the MAB or CogScreen may be taken as evidence of an impairment.

Level of Performance Method To date, only a very small percentage of USAF aviators have archived EFS testing. As such, methodologies are necessary for the assessment of aviators without pre-morbid testing. Here the EFS data on MAB and CogScreen variables may be used as a group reference. Pilots with poor performance on testing following some insult may be inferred to be at that low level of performance due to the cortical insult. Aviators who are found to be in the bottom one percent following some trauma, for example, are statistically more likely to be at that level due to the trauma than due to their initial performance. In other words, there would only be a one percent chance that the aviator was pre-morbidly at that low level of performance.

In order to effectively utilize this approach, a number of statistics and tables are necessary. First, the means and standard deviations of a large sample of fairly similar individuals is required. This provides the norm against which to compare a new individual's scores. In addition to these statistics, percentile levels of various scores are often of use. While the mean and standard deviations model the underlying distribution of test scores

when the distribution is normal, they do not model skewed distributions well where there is an asymmetry in scores. Providing the scores of a distribution at critical percentile points allows the scores of new patients to be very accurately placed relative to their peers.

<u>Pattern of Performance Method</u> While the above method uses a large group of subjects as the comparison for an individual's post-insult scores, it is also possible to use some elements of the person's own performance to make conclusion regarding cognitive change. A common approach uses the effects of aging on various types of test performance as a model. It has long been known that some types of intellectual ability are fairly sensitive to aging and other types are quite resistant to change. Classically, these are referred to as "hold" and "don't hold" variables. Scores on tasks such as vocabulary and general information generally are similar across age brackets. These tasks tend to "hold" as one ages. Scores on other tasks such as performance type tests like speeded, visuo-motor ability usually drop off with age. Here, somewhere in the fifth decade of life, performances "don't hold" and begin a fairly constant decline.

Applying this method to younger patients who have had some type of cortical insult suggests that larger differences in scores between "hold" and "don't hold" tests is associated with greater levels of impairment. It is common, for example, to look at the difference between the Vocabulary subtest on the WAIS-R and the Digit Symbol subtest. If the Digit Symbol subtest is more than 2 or 3 standard scores below the Vocabulary subtest score (and there is history of

insult), there is a good likelihood of impairment.

There are always naturally occurring differences between two sub-tests on any test. It is, therefore, necessary to quantify this natural difference so that referred aviators might be compared to the "normal" differences. Aviators whose difference scores between two tests are in the top 99% of non-impaired aviators can be assumed to have that level of difference due to insult, as the a priori chance of that difference is quite low.

Results and Application

Change in Performance Method

Table 1 (all tables are available from the author due to page constraints here) provides the means and standard deviations for each of the MAB scores. These include summary scores as well as scaled and raw scores. The scaled scores are based upon the 1990 norms. The raw scores are provided here and in subsequent tables in the event that there is a re-norming of the test. As can be seen, pilots are on average quite intelligent with Full Scale IQ scores of 119. This table also includes the stability coefficient, the standard error of measurement, and the 95% confidence band for each of the scores. The stability coefficient is based upon the testing and retesting of a group of subjects during the development of the test. It indicates the degree to which scores remain constant across time. The standard error of estimate statistic indicates the variability of scores that could be expected from multiple testings of the same person. Finally, the 95% confidence band indicates the differences in scores that might be expected at the 95% probability level. This final confidence band can be applied to any individual's scores. If a second testing is

below the confidence band, the performance should be interpreted as lower and more deficient than what can be expected simply due to measurement error.

As an example, suppose a pilot received a Full Scale IQ score of 125 during initial EFS screening. The pilot is then involved in a car accident with a brief coma. The pilot is referred for follow-up cognitive testing. The expected range of scores for this pilot would be 125 plus or minus 2.38 points. As such, the range would be 123 to 127. The MAB is re-administered and the Full Scale IQ score is 118. Since this is well below the bottom of the confidence band (123), there is good reason to suspect a true decrement in ability. Obviously, it is another question whether an IQ of 118 is too low to continue flying; nevertheless, an impairment is verified. Other testing and other evidence can go to the question of continued flying.

There are ten subscales which can also be used to answer more specific functional questions in the same manner. It is of particular importance when a referral question specifically mentions an error of concern such as spatial ability and subsequent testing indicates performance on the spatial subtest well below the confidence band. Additional evidence might be gathered from the number of subscales below the bands. A pilot with only one of the tests below the band is very different from a pilot with all ten subtests below the bands.

With 65 variables, the CogScreen is somewhat difficult to interpret (See Appendix A for variable names). In order to better understand the data, it is presented not by subtest but by type of score. As such, speed variables are presented first, followed by

accuracy, throughput, and process variables.

Table 2 provides not only the means and standard deviations for the CogScreen speed variables but also the stability coefficient, the standard error of estimate, and the 95% confidence band. The stability coefficient was taken from the test manual and used specifically to develop the other two statistics for this sample.

Here, for example, a subject's reaction time speed score on the Math task would have to be banded by plus and minus 8.26 seconds. As such, a subject with a pre-morbid score of 30.00 seconds would have a 95% statistical probability of producing a score between 21.74 and 38.26 seconds. A clinically important finding would be a score significantly slower such as 42 seconds. In this example, a pilot with a pre-morbid score of 30 seconds probably has a decline from prior functioning with that score of 42 seconds. Conversely, a post-morbid score of 35 seconds is within the measurement error range and should not be clinically interpreted as a decline.

With so many speed scores, it is important not to calculate so many statistics on a single patient that the method becomes a "fishing trip" with a "drift net". The two tasks with the best speed characteristics are probably the MTS (Matching to Sample) and MAN (Manikin). These tasks require a small amount of cognitive performance directed toward a fairly focal stimuli. With average performance in the one and a half to two second range, there is sufficient room for variable performance. Tasks which have much shorter reaction times are probably prone to be confounded by the use of the light pen, the use of large muscle groups, subtle shifts of

position, administration differences, and software changes. Tasks such as MATH are not true reaction times. The 30 seconds or so of task time includes attention, reading speed, math calculation time, and reaction time. As such, it is a heterogeneous task, and hence of limited interpretive value here.

Tables 3, 4, and 5 provide the means and standard deviations for the accuracy, throughput, and process variables. The accuracy scores have so little variance in normal pilots that the calculation of stability coefficients, standard errors of measurement/ estimate, and confidence bands is inappropriate. This lack of variance is also noted in the manual for the normative sample. The reason that the scores vary so little is due to "ceiling effect". The tasks are so easy that most subjects (at times over 90%) get all tasks correct and as such there is no separation of performance on the high end of ability. Since throughtput variables are the product of speed and accuracy variables, they add little information over the speed data. Finally, the manual does not present stability data for the process variables and as such confidence bands cannot be calculated. Here is an example of where a USAF stability study would allow for such data.

<u>Level of Performance Method</u>

Table 6 provides the percentile levels for the MAB variable distributions. A subject with a score of 129 would be at the 95% and be quite intelligent compared to other pilots. For clinical purposes with a patient who did not have prior testing, these data can be interpreted as the probability of a post-insult decrement in functioning.

The chances that a pilot has a Full Scale IQ score of 100 is about 1%, because only 1% of the sample have Full Scale IQ scores of 100 or less. One way to interpret this data clinically is to say that there is a 99% chance that the pilot with the IQ of 100 had an IQ of greater than 100 prior to any cognitive insult. Here the very fact of exceptionally low performance is in and of itself unlikely and most probably due to clinical factors.

In general, scores in the lower 1% and 5% levels are probably clinically relevant. Again, the quality (which scales) and quantity (how many scales) are of interest. Performance scores and tasks are more important for aviators and also more prone to cognitive decline with insult. Conversely, pilots with scores in the top 95% and 99% are probably able to return to duty and clinical significant impact is highly unlikely.

Tables 7, 8, 9, and 10 provide similar cutscores for the CogScreen speed, accuracy, throughput, and process distributions. For the speed data in Table 7, performance is in seconds and therefore larger numbers represent poorer performance. While on the MAB higher scores are better, here lower scores are better. Very fast answering of the Math items might result in a score of 15 seconds. This would place that subject at the 5% level, a very good performance.

A patient, however, who spends 45 seconds on average would be somewhere between the 95% and 99% level. That patient had a very small chance of taking that much time given the group norms and so is probably impaired. Again, the quality and quantity of scores must be part of the clinical decision process.

Again, as with the speed variables in the CogScreen. It is recommended that the Matching to Sample and Manikin tasks be used for most clinical work. They exhibit good range across the sample and are less prone to error than the faster, pure reaction time tasks.

Table 8 provides the tail of the distribution associated with low accuracy scores. Full tables are not possible due to the limited variance of these scores. In essence, most pilots got these tasks right with a few pilots getting some tasks wrong. Using Math as the example again, a pilot who only gets a .20 proportion of the Math questions correct is at the bottom 1% of the distribution. A .40 proportion would place that pilot at only the 15% level. Either score should be of clinical concern.

Table 9 presents the throughput data. Here, higher scores represent very fast, accurate, and efficient cognitive processes. Low scores represent poor performance. A throughput of 0.3 on the Math task would represent a performance at the first percentile of the distribution. This would suggest a impairment relative to the norms.

Finally, Table 10 presents the distributions for the process variables. The table's footnote indicates the direction of performance and the tails of clinical concern. Here, again, a number of the variables had highly skewed distributions with limited variance and only a limited number of distribution points could be mapped.

<u>Pattern of Performance Method</u>

Table 11 provides the statistically expected differences in scaled scores across tests given to a single subject at a single point

in time. The MAB is used here because the variables are widely used and understood. The CogScreen is not presented because no theory or research exists on its interscale behavior in impaired individuals.

The approach here is that variables such as Vocabulary and Information are relatively resistant to cognitive insult. The performance tasks (Digit Symbol, Picture Completion, Spatial, Picture Arrangement, and Object Assembly) are far more likely to be affected by an impairing incident. Difference scores, however, will naturally vary quite widely in non-impaired individuals and must be modeled.

To develop this data, the scaled scores for each of the performance tasks was subtracted from the scaled score of Vocabulary and Information. This resulted in a distribution of difference scores for the sample. The means and standard deviations are presented in Table 11. On average, pilots have better performance scores than Vocabulary scores as evidenced by the negative difference scores. Their scores on Information are more similar to, and slightly better than, their scores on the performance tasks with difference scores of generally 1 to 3 points.

The data of interest are those differences which are positive and large. This would clinically suggest that performance type ability is well below the traditional "hold" verbal tests. The "hold" tests would have "held" and the "don't hold" tests would have "not held". The bottom line of a positive and large score would be a cognitive impairment.

If a patient had a Vocabulary score of 60 and a Digit Symbol

score of 45, the difference would be 15 points. Looking at the table, a 15 point difference would place this patient well above the 99th percentile. A clinician could be 99% certain that such scores would not be found in non-impaired pilots.

It is recommended that the Scaled Vocabulary score minus the Scaled Digit Symbol score be used for most purposes. Vocabulary seems to behave best in this population and appears to have the most stable norming across studies using the MAB. Digit Symbol is a complex, heterogeneous task which is sensitive to many functional declines. The raw score difference scores are unstable due to the lack of a common underlying metric and are provided here for reference only.

DISCUSSION

The accurate assessment of the cognitive functioning of pilots is essential. The lives and careers of pilots and the lives of crews and passengers may depend upon it. The USAF also is interested in increasing mission effectiveness, reducing training costs, and managing retention.

The USAF Enhanced Flight Screening program has provided an opportunity to collect large sets of cognitive data on pilot candidates. No other study or function has ever allowed for such large samples or for the archiving of individual data.

Three clinical methods for the neuropsychological assessment of pilots have been delineated. A method using pre-morbid test data for those pilots with archived EFS data has been explored. Additionally, two methods have been explained for the testing of pilots without pre-morbid testing available. The necessary statistical tables are

presented for clinical use.

A number of caveats must be mentioned. First, these data are from pilot candidates. Not student pilots and not pilots. As such there is some chance that the data are not as precise as they might be. A number of studies, however, have found very similar intelligence test data. Also Retzlaff, King, and Callister (1995) found no differences in intelligence between those entering pilot training and those finishing. The CogScreen is less well known and larger differences may operate.

It would also have been better to use stability coefficients which had been calculated from an Air Force pilot sample. The use of general stability coefficients from the test manuals are within the normal range of practice, but a one year test-retest study of a group of mid-career pilots would have provided much more specific statistics.

Finally, it is important to note that this is a relatively atypical approach to neuropsychology driven by the unique needs of the USAF medical baselining requirements. Psychology has a long history of neuropsychological tests, assessment, and methods. Traditional neuropsychological assessment includes many tests across many hours of individualized testing. It is fully expected that the current work will be in addition to, not in place of, the traditional techniques.

REFERENCES

Callister, J. D., King, R. E., & Retzlaff, P. (1995). Cognitive assessment of USAF pilot training candidates: Multidimensional Aptitude Battery and CogScreen Aeromedical Edition. (AL/AO-TR-1995-0125). Brooks AFB: Armstrong Laboratory.

Flynn, C. F., Sipes, W. E., Grosenbach, M. J., and Ellsworth, J. (1994). Top performer survey: Computerized psychological assessment of aircrew. Aviation, Space, and Environmental Medicine, 65, 39-44.

Jackson, D. N. (1985). Multidimensional Aptitude Battery. Port Huron, MI: Research Psychologists Press.

Kay, G. G. (1995). CogScreen-Aeromedical Edition Professional Manual. Odessa, FL: Psychological Assessment Resources, Inc.

King, R. E. and Flynn, C. F. (1995). Defining and measuring the "Right Stuff": Neuropsychiatrically Enhanced Flight Screening (N-EFS). Aviation, Space, and Environmental Medicine, 66, 951-956.

Retzlaff, P. and Gibertini, M. (1988). The objective psychological testing of Air Force officers in pilot training. Aviation, Space, and Environmental Medicine, 59, 661-663.

Retzlaff, P. and Gibertini M. (1987). Air Force pilot personality: Hard data on "The Right Stuff". Multivariate Behavioral Research, 22, 383-399.

Retzlaff, P. & Gibertini, M. (1994). Neuropsychometric issues and problems. In Vanderploeg, R. (ed.), Clinician's Guide to Neuropsychological Assessment. Hillsdale, NJ: Erlbaum.

Retzlaff, P., King, R. E., & Callister, J. D. (1995). Comparison of a computerized version to a paper/ pencil version of the Multidimensional Aptitude Battery. (AL/AO-

TR-1995-0121). Brooks AFB: Armstrong Laboratory.

Retzlaff, P., King, R. E., & Callister, J. D. (1995). US Air Force pilot training completion and retention: A ten year follow-up on psychological testing. (AL/AO-TR-1995-0124). Brooks AFB: Armstrong Laboratory.

Vanderploeg, R. (ed.) (1994). Clinician's Guide to Neuropsychological Assessment. Hillsdale, NJ: Erlbaum.

THE USE OF SOLID-PHASE MICROEXTRACTION (SPME) FOR THE LOW LEVEL DETECTION OF BTEX AND PAHs IN AQUEOUS LEACHATES

William G. Rixey
Assistant Professor
Department of Civil and Environmental Engineering

University of Houston
4800 Calhoun
Houston, TX 77204-4791

Final Report for:
Summer Faculty Research Program
Armstrong Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, DC

and
Armstrong Laboratory

October 1996

THE USE OF SOLID-PHASE MICROEXTRACTION (SPME) FOR THE LOW LEVEL DETECTION OF BTEX AND PAHs IN AQUEOUS LEACHATES

William G. Rixey
Assistant Professor
Department of Civil and Environmental Engineering
University of Houston

Abstract

The use of Solid-Phase Microextraction (SPME) was investigated for the low level detection of BTEX and PAH compounds in aqueous leachates. Equilibrium partition coefficients and rates of sorption to the SPME fiber from aqueous solutions of the contaminants were determined. Equilibrium results compared favorably with literature data and demonstrate that the technique is particularly sensitive for the low level detection of PAH compounds, e.g., phenanthrene. Direct liquid phase sorption to the fibers was compared with sorption in the head-space above the aqueous solution. Results indicate that direct liquid phase sorption results in shorter equilibration times and in less interference for both laboratory prepared aqueous solutions as well as field leachates from complex fuel mixtures.

THE USE OF SOLID-PHASE MICROEXTRACTION (SPME) FOR THE LOW LEVEL DETECTION OF BTEX AND PAHs IN AQUEOUS LEACHATES

William G. Rixey

I. Introduction

This objective of this research was to develop a method for measuring BTEX and PAH compounds in aqueous solutions by the solid phase microextraction (SPME) technique. This SPME method is desirable for PAH compounds because lower detection levels in aqueous samples are possible than those for solvent extraction methods. In addition the technique will also permit the analysis of both BTX and PAHs in a single extraction/GC analysis.

The SPME method is being developed for two applications: (1) as an analytical tool for assessing the long-term aqueous leaching characteristics of BTEX and PAH compounds from contaminated soils and oily wastes and (2) as method for determining BTEX, and low level PAH, and TPH concentrations in field groundwater samples and in leachates from field soils. In addition to BTEX and PAH analysis the SPME method is also potentially useful for full-range TPH in groundwater analysis with a single step extraction/GC procedure.

Specific objectives of research reported here were as follows:

1) obtain partition coefficients for SPME fiber for BTX and Phenanthrene and compare with the limited measurements reported in the literature.

2) measure rates of sorption to determine required equilibration times for commercially available fibers.

3) compare methods of sorbing contaminants from aqueous solution - headspace vs. direct liquid phase sorption.

4) determine reasons for interference with semi-volatile range compounds and evaluate a methodology for minimizing this interference.

II. Background:

The SPME technique was developed at the University of Waterloo (1,2) and eliminates many drawbacks with solvent-based extraction methods for contaminants in aqueous samples. Recently, it has been commercialized for use in detecting both volatile and semi-volatile compounds at low levels in water (3).

III. Materials:

SPME Fiber:

Two 95 μm Supelco SPME fibers and a manual syringe holder for the SPME fibers were obtained for these studies. The SPME fiber had the following dimensions:

O.D of polymer coating: 300 microns

I.D. of polymer coating: 110 microns

Length of polymer coating: 1 cm

Polymer material: polydimethylsiloxane

Contaminants:

For preliminary studies to determine the equilibrium partition coefficients and the rates of sorption of various contaminants, benzene, toluene, ethylbenzene, m&p xylene, o-xylene, and phenathrene were used. The initial concentrations of the aqueous solutions were 26 μg/L for benzene, toluene, ethylbenzene, m- & p-xylene, o-xylene, and 74 μg/L for phenanthrene. These solutions were prepared by first dissolving_the contaminants in methanol and then adding the methanol solutions to Nanopure ® water.

IV. Methods:

Sorption properties and rates of sorption were determined by exposing a SPME fiber from a SPME syringe to aqueous solutions of dissolved contaminant. The sorption to the fiber was carried out in two ways:

(1) SPME fiber placed in headspace above an aqueous sample

(2) SPME fiber placed directly in aqueous sample.

For each case the aqueous solutions were stirred continuously at a constant speed for each sorption experiment. Sorption was carried out in EPA glass vials (nominally 40 ml) with teflon septa. The amount of aqueous solution used was the same for each of the two types of sorption experiments: headspace or direct aqueous SPME sorption.

After sorption for a various times, the fiber was retracted into the syringe, the syringe removed from the teflon septum, and then the syringe was injected into a gas chromatograph with FID detector. The GC conditions were as follows:

GC Conditions:

Gas Chromatograph:

Hewlett Packard 5890 GC with FID Detector

GC Program:

1) 35C hold for 6 minutes then ramp at 1 C/min up to 45C

2) Ramp 40C/min to 280C, hold for 10 min.`

Detector temp: 280 °C; Injector temp 250 °C

Splitless injection with purge valve on at 0.5 min

Amount injected: 1 μL

Total run time: 31.87 minutes

Column: DB-1 15 meters by 0.25mm, 1μm film thickness

Spectra Physics PC-1000 Data Station (PC1000 ver 2.5)

The SPME fiber remained in the injector throughout the run in order to effectively desorb the contaminants from the fiber for repeated sorptions and GC analyses. It was found that some carryover was still observed for the higher boiling contaminants such as phenanthrene, and that onger desorption times may be required for very low level detection of PAHs in water. Our 74 μg/L phenanthrene solutions, the impact of carryover on detection was negligible.

Sample chromatograms is shown as Figures 3 and 4.

V. Equilibrium Partition Coefficients for BTEX and PAHs to a SPME Fiber:

A mass balance at any time t yields in the aqueous phase:

$$C^w = C_o^w - \frac{V^p}{V^w} C^p \qquad (1)$$

where:

C^w	=	concentration of contaminant in the aqueous phase at any time, t (g/cm^3 w)
C_o^w	=	concentration of contaminant in aqueous phase initially (g/cm^3 w)
C^p	=	concentration of contaminant in the polymer film at any time, t (g/cm^3 polymer)

Similarly, in terms of the aqueous concentrations, the concentration in the polymer at any time t is:

$$C^p = \frac{V^w}{V^p}(C_o^w - C^w) \qquad (2)$$

As the concentrations in the polymer and aqueous phases approach the equilibrium concentrations, i.e., $Cw \rightarrow C_\infty^w$ and $Cp \rightarrow C_\infty^p$ and they will be related by the linear, equilibrium partition coefficient, K_p given by.

$$K_p = \frac{C^p}{C^w} \qquad (3)$$

Thus, the polymer concentration at time $t \rightarrow \infty$, in terms of the initial aqueous concentration, becomes:

$$C_\infty^p = \frac{C_o^w}{\frac{1}{K_p} + \frac{V^p}{V^w}} = \frac{K_p C_o^w}{1 + \frac{K_p V^p}{V^w}} \qquad (4)$$

We will use this equation when looking at the time to approach equilibrium in the next section.

An important considerations is the amount of material that is predicted to be removed from the sample, we will refer to this as fractional recovery or fractional sorption of the original solute in the sample, *f*, where $f = 1 - C_\infty^w / C_o^w$.

$$f = 1 - \frac{C_\infty^w}{C_o^w} = 1 - \frac{1}{1 + \frac{K_p V^p}{V^w}} = \frac{\frac{K_p V^p}{V^w}}{1 + \frac{K_p V^p}{V^w}} \qquad (5)$$

Values of f for various values of K_p calculated from Equation 5 are shown in Figure 1.

The following values were used to in calculations of V_p and V_w for calculations using Equation 5:

Dimensions of the polymer:

Length of exposed fiber: 1 cm,

O.D: 300 microns or 0.030 cm

I.D.: 110 microns or 0.011 cm

Volume of polymer phase: 0.000612 cm^3

Volume of aqueous phase: 35 cm^3

therefore, $V^p/V^w = 1.74 \times 10^{-5}$ (cm^3-polymer)/(cm^3-w).

Figure 1 shows that the sensitivity of the SPME technique increases with increasing partition coefficient. These calculations indicate that more than 10% of the solute in a 40 ml aqueous sample will be sorbed (assuming equilibrium conditions) for solutes with K_ps greater than 10000. Thus, contaminants such as phenanthrene (Kow>10000) will be potentially measurable at very low detection limits by this technique, i.e., as low as 1 µg/L by capillary GC with a FID detector assuming a 40 ml sample and no interference by other compounds with similar retention times. See Section IX for a discussion about possible interferences.

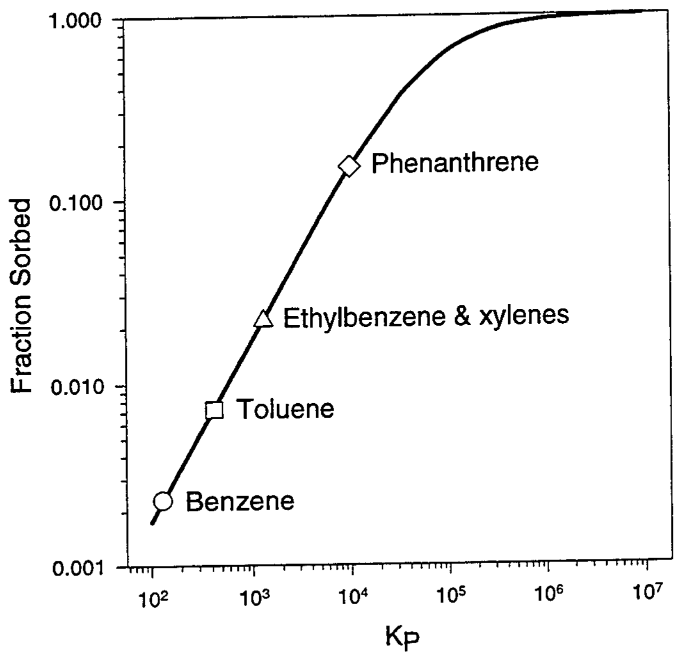


Figure 1. *Fraction of contaminant sorbed to an SPME Fiber. Curves are calculated curves based on equilibrium partitioning from 35 ml of an aqueous phase to a 95 μm phase supported by a fiber with an outer diameter of 300 μm. The points represent measured K_ps for benzene, toluene, ethylbenzene & xylenes, and phenanthrene.*

The measured K_p is calculated as:

$$K_p = \frac{V_w}{V_p}\frac{f}{1-f} \qquad (6)$$

VI. Experimental Equilibrium Partition Coefficients for BTEX to a SPME Fiber:

The partition coefficients for BTEX were determined from the fractional uptakes of BTEX present in 35 ml of aqueous solution and the use of Equation 6. Fractional sorptions for BTEX and phenanthrene are shown in Table 1. Experiments were conducted for two sorption times for BTEX and three sorption times for phenanthrene, each starting with a 35 ml aqueous sample containing 26 μg/L each of benzene, toluene, ethylbenzene, m-xylene, p-xylene, and o-xylene, and 74 μg/L phenanthrene. The fractional uptake was determined from the uptake on the fiber measured by direct

injection into a GC. Note that the fractional uptake did not vary significantly with time for the BTEX components indicating that for the BTEX components equilibrium sorption had been reached.. (This conclusion is consistent with analysis of rate of sorption of phenanthrene onto the fiber. See Figure 2 and the discussion in Section VIII). The results demonstrate that BTEX sorption on even the 95 μm fiber reaches equilibrium in less than 30 minutes. (It is further shown in the next section and in Figure 2 for contaminants with Kps less than 1000, that equilibrium soprtion conditions should be reached within 30 minutes. For contaminants with Kp >1000, longer sorption times are required to reach equilibrium sorption.)

Note that the fractional uptake for benzene is only 0.23% when extracting benzene from a 35 ml aqueous sample with a 95 μm SPME fiber. This is a result of the relatively low K_p for benzene. Fractional recovery increases with partition coefficient as shown in Figure 1 and also in the data of Table 1. Xylenes have nearly 10 times the equilibrium sorption capacity as that for benzene. Table 1 shows that more than 2% of the xylenes in the 35 ml sample are sorbed to the fiber at equilibrium. This SPME technique will be particular sensitive to the higher molecular weight monoaromatic compounds, e.g., xylenes and $C3^+$- alkyl aromatics and polyaromatic hydrocarbons, e.g., phenanthrene, fluorene, etc.

The measured partition coefficients, K_p. are compared with K_{ow}s and measured K_ps for SPME fibers reported in the literature. The values for K_p compare favorably with the exception of the literature K_p values for ethylbenzene and o-xylene. The value for ethylbenzene for this study, however, is more consistent with the K_{ow} value.

VII. Experimental Equilibrium Partition Coefficients for Phenanthrene to a SPME Fiber:

In contrast to the results for BTEX, the equilibrium K_p for phenanthrene could not be determined from the headspace analysis results shown in Table 1. The sorption of phenanthrene did not reach equilibrium when the SPME fiber was placed in the headspace of the vial containing the aqueous solution. The 20 hr sorption measurement was not shown for phenanthrene because there it was believed that there was significant interference with a peak desorbing from the vial septum. As result, subsequent sorption

experiments were conducted with the SPME fiber in direct contact with the aqueous sample. This resulted in approximately four times faster sorption of the phenanthrene from aqueous solution. Sorption was conducted for three times, 30 min, 65 minutes, and 2 hours. The sorption at two hours had not quite reached equilibrium, so the fraction removed at equilibrium cannot be directly determined form the values in Table 2. However, fitting the fraction removal data of Table 2 to an appropriate rate equation will yield both equilibrium sorption constants as well as sorption rate constants as demonstrated in Section VIII. From this analysis of the data of Table 2 the equilibrium sorbed fraction onto the SPME fiber was determined to be 0.15 for phenanthrene. From this value of f=0.15 a K_p=10,000 was calculated which compares favorably with a literature value of 13,500 for anthracene.

VIII. Rates of Sorption for Various Contaminants to a SPME Fiber:

Consider that the mass transfer is limited by diffusion in the liquid film surrounding the fiber. This is probably more likely for the solutes with higher K_{ow} since permeabilities will be higher in the polymer film as the diffusivity in the polymer phase does not decrease proportionately with increasing K_{ow}

Again a mass balance at any time yields:

$$\pi(R_o^2 - R_i^2)L\frac{dC^p}{dt} = 2\pi R_o Lk(C^w - \frac{C^p}{K_p}) \qquad (7)$$

where

k	=	mass transfer coefficient in aqueous liquid film, cm/sec
L	=	length of SPME coating on fiber, cm
R_o	=	Outer radius of coating, cm
R_i	=	Inner radius of coating, cm

Let's assume that we develop a film resistance to mass transport: The mass transfer coefficient will be given approximately by D/R for a cylinder.

We can substitute the expression for C_w in terms of C_p from the mass balance at any time, t given by Equations 1. After substituting and integrating we obtain:

Table 1. Fraction BTEX and Phenanthrene Sorbed to a SPME fiber for various exposure times. SPME fiber placed in Headspace during Sorption from the Aqueous sample.

Contaminant	30 minute Sorption	50 minute Sorption	20 hr Sorption
Benzene	0.0023	0.0023	-
Toluene	0.0074	0.0071	-
Ethyl benzene	0.022	0.022	-
m and p xylene	0.022	0.022	-
o-xylene	0.019	0.020	-
phenanthrene	0.014	0.028	**

** - measurements at long times were affected by interference with other compounds. See Section IX on compound interference.

Table 2. Fraction BTEX and Phenanthrene Sorbed to a SPME fiber for various exposure times. SPME fiber placed directly into the Aqueous Phase during Sorption from the Aqueous sample.

Contaminant	30 min Sorption	65 minute Sorption	2 hr Sorption
Benzene	-	-	-
Toluene	-	-	-
Ethyl benzene	-	-	-
m and p xylene	-	-	-
o-xylene	-	-	-
phenanthrene	0.055	0.105	0.125

From these measurements an equilibrium sorbed fraction was determined to be 0.15 from a best fit of Equation 9 to the data. See Figure 2 and Section VIII on rates of sorption to SPME fibers.

Table 3. Fraction BTEX and Phenanthrene Sorbed to a SPME fiber for various exposure times. SPME fiber placed directly into the Aqueous Phase during Sorption from the Aqueous sample

Contaminant	K_{ow} (Literature) (g_i/cm^3 octanol/ g_i/cm^3 w)	K_p Measured (1) (g_i/cm^3 polymer/ g_i/cm^3 w)	K_p (Literature) (g_i/cm^3 polymer/ g_i/cm^3 w)
Benzene	135	130	126(2)
Toluene	490	420	340(2)
Ethyl benzene	1413	1300	528(2)
m and p xylene	1499	1300	-
o-xylene	589	1150	654(2)
phenanthrene	37154	10000	13500(3)

(1) this study - 95 μm polydimethylsiloxane SPME film.
(2) measurements onto a 56 μm methyl silicone film [Arthur et al., 1992].
(3) estimated from measurements for anthracene onto a SPME fiber (15 μm polydimethysiloxane) [Potter and Pawliszyn, 1994].

$$C^p = \frac{K_p C_o^w}{1 + \frac{K_p V^p}{V^w}} \left(1 - \exp\left[- \frac{2R_o}{R_o^2 - R_i^2} \frac{k}{K_p} (1 + \frac{K_p V^p}{V^w}) t \right] \right) \quad (8)$$

The above equation can also be written in terms of approach to equilibrium:

$$C^p = C_\infty^p \left(1 - \exp\left[- \frac{2R_o}{R_o^2 - R_i^2} \frac{k}{K_p} (1 + \frac{K_p V^p}{V^w}) t \right] \right) \quad (9)$$

Equation 3 shows that it will take longer for solutes with larger values of K_p to reach equilibrium. Equation 3 can be used to estimate the time required for equilibrium for solutes. Note that the time to reach equilibrium becomes independent of solute K_{ow} for high values of K_{ow}. This is because the amount of mass that is transferred no longer is proportional to the K_{ow}, i.e., at high K_{ow}s we have essentially transferred a finite amount of mass.

IX. Potential Interference:

Interference with some the compounds was observed when the SPME fiber was placed in the headspace of the aqueous samples. Table 4 shows the typical unknown peak retention times and areas which were observed in samples containing the BTX and phenanthrene standards. These are the peaks with retention times of 18.58, 19.65, 20.47, 21.16, 22.23, and 24.12 min. (For comparison the retention time for phenanthrene was 22.58 minutes for our GC temperature program and an area of 1150000 was obtained for a 2 hr exposure using the SPME fiber in the aqueous phase with a 35 ml aqueous sample containing 74 μg/L phenanthrene). These interfering peaks are significantly reduced when the fiber is exposed directly with the aqueous phase.

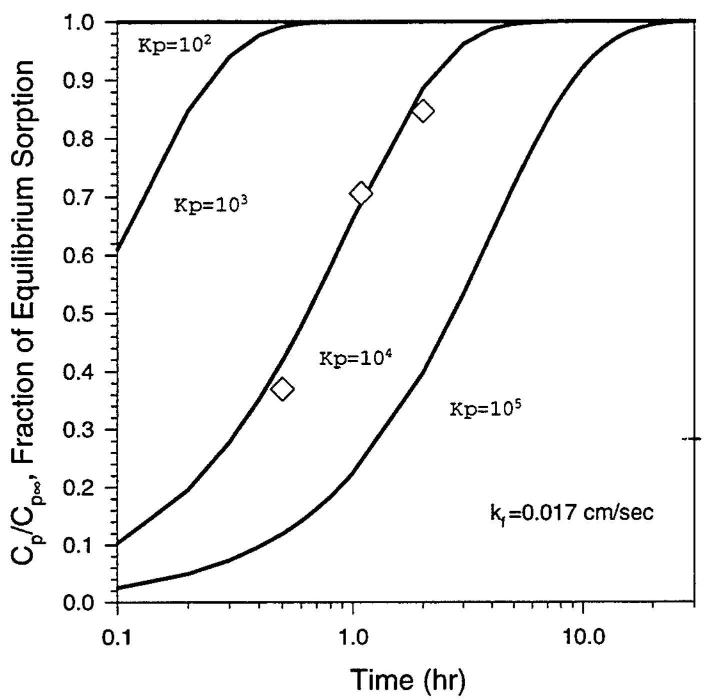


Figure 2. *Sorption onto a 95 μm SPME fiber from a 35 ml aqueous sample as a function of time for various equilibrium partition coefficients. Curves were fit to data for phenanthrene. The data points were taken from those in Table 2. The best fit yielded a $K_p=10^4$ (f=0.15) for phenanthrene and a film coefficient $k_f=0.017$ cm/sec. It has been assumed that film diffusion is controlling the mass transfer.*

Table 4. Effect of SPME Exposure Time on Interference of GC peaks with PAH Peaks. GC peak areas are listed. SPME fiber was exposed to head space of a clean (new) 40 ml EPA vial with 35 ml of nanopure water.

Retention Time, min	2 hrs	14 hrs	41 hrs
18.58	36135	132530	120509
19.65	148040	364221	556502
20.47	102907	561782	1091835
21.16	52455	298333	563163
22.23	72349	290484	772921
24.12	137391	136845	151589

Table 4 shows the increase in area for the other peaks as a function of time. It was suspected that there may be contamination from the fibers, contamination from the room due to exposure of the fibers between injections, or contamination from the DI water used to make the aqueous samples, or contamination from the clean vials themselves. A series of experiments was carried out to determine the effect of each of these parameters.

1) Sorption from the atmosphere.

It was considered that there might be material from the room that was sorbing and contributing to the interference of the peaks. As a result an experiment was conducted such that the SPME fiber/syringe was placed in the room for sorption immediately after the contaminants were desorbed in the injector port of the GC at 250 °C for two hours. The results showed that subsequent injections showed no visible peaks interfering with the PAHs.

2) Sorption from aqueous samples:

To test whether there was an effect of contamination present in the aqueous samples themselves, "blank" samples of water were sorbed onto the fibers and injected under the

same conditions as the other samples. It was observed that the same peaks were observed.

3) Sorption from "clean" vials:
SPME fibers desorbed overnight at 250 °C were placed in "Clean" EPA vials which were factory sealed prior to use. The syringe was injected into the vials containing no aqueous solution - only air in the vial - and the same unknown peaks were observed, thereby indicating that the vial septa contain semivolatile compounds which can interfere with analysis of PAH compounds using SPME with headspace sorption. However, as noted previously this interference can be reduced by sorbing directly from the aqueous phase.

X. Analysis of Field Samples:

The two methods of SPME sorption, headspace vs. direct injection, were used on groundwater sample which had been contaminated by petroleum hydrocarbons at the FT-23 Active Fire Training Area, Tyndall Air Force Base, Florida. Chromatograms for samples taken from monitoring well - MWQ1 are shown in Figures 3 and 4. Figure 3 is a chromatogram for headspace sorption onto the SPME fiber for 1 hr and Figure 4 is a chromatogram for direct aqueous sorption onto the SPME for 1 hr. These results show that there is a significant response for the fiber to semivolatile range compounds. The peaks in the semivolatile range are most likely due to the other hydrocarbons that are present in the fuel mixture that come on contact with the groundwater (for the short exposure the contributions from the septum to the headspace sorption should be relatively small). Note that a significant number of the peaks in the semivolatile range are significantly reduced when direct aqueous sorption is used. The likely cause for the difference is that a film of hydrocarbon is present and that semivolatile alkanes and alkenes have such low solubilities that they do not have enough time to sorb to the fiber from through the aqueous phase. It would be expected that if longer times were used for the sorption then the responses would be expected to be the same.

These results do indicate however that the SPME technique, especially direct aqueous sorption, has considerable potential for BTEX and semivolatile analysis for field contaminated aqueous samples.

XI. Conclusions:

1) Volatile compounds such as BTEX can be extracted efficiently from the aqueous phase by extracting with SPME fibers placed in the head space above the aqueous sample. However, head space analysis should not be used for semi-volatile compounds. Longer sorption times are required for headspace analysis - sorption times were observed to be 4 times longer by headspace analysis vs. direct liquid phase sorption to SPME for phenanthrene. Moreover, when analyzing the headpace there is significantly more interference with PAHs from other compounds in the sample as well as contaminants that are present from other sources, e.g., vial septa.

2) Liquid phase analysis results in less interference. In addition to interference from other contaminants present in an aqueous sample, interference can be caused by contaminants present in aqueous vials. These contaminants apparently diffuse from the teflon septum into the vial headspace. These contaminants are apparently not very water soluble - since interference by these contaminants in the aqueous phase was significantly reduced.

3) Equilibration times of three hours are estimated for phenanthrene ($Kp=10^4$) for sorption from a 35 ml aqueous sample with a 95 μm SPME fiber. Shorter equilibration times could be used, but less material will be sorbed. Alternatively, a thinner film of material could be also be used. This would insure that samples will reach equilibrium which would avoid having to carry out sorption at fixed time intervals for compounds which do not reach equilibrium for thicker SPME films. For a 95 μm fiber equilibration times of 24 hours would be required for analysis of the entire TPH range which would include contaminants with K_ps > 10^5 . Analysis of TPH in groundwater by SPME therefore could be conducted with two fibers - a 95 μm fiber for contaminants with K_{ow}

<1000 and a 7 μm fiber for contaminants with K_{ow}>1000 to maximize sensitivity with a practical, 30 minute sorption time.

4) Rates of sorption appear to be aqueous film controlled. Measurements of the rate of sorption suggested a reasonable film coefficient for the experimental conditions for these measurements.

XII. References:

1) Arthur, C. L., et al., "Analysis of Substituted Benzene Compounds in Groundwater Using Solid-Phase Microextraction," *Environ. Sci. Technol.*, **26**, 979 (1992).

2) Potter, D. W., and J. Pawliszyn, "Rapid Determination of Polyaromatic Hydrocarbons and Polychlorinated Biphenyls in Water Using Solid-Phase Microextraction and GC/MS," *Environ. Sci. Technol.*, **28**, 298 (1994).

3) Solid Phase Microextraction of Semivolatile Compounds in US EPA Method 625, Supelco Application Note 6, Supelco, Inc., 1994.

Name: SPME-MWQ1
Description: SPME-MWQ1-1 hr
Type: Sample
Injection Volume: 1.0 uL

Vial: 1

Injection: 1 of 1

Injected On: 07-26-96 11:09:17

Acquisition Log
Column Pressure: N/A
Noise (microAU): 3
Run-Time Messages: None

Column Temperature (C): N/A
Drift (microAU/min): 0.06

Pump Flow Stability: N/A

Signal 1: Interface A
Calculation Type: External Standard (Area)

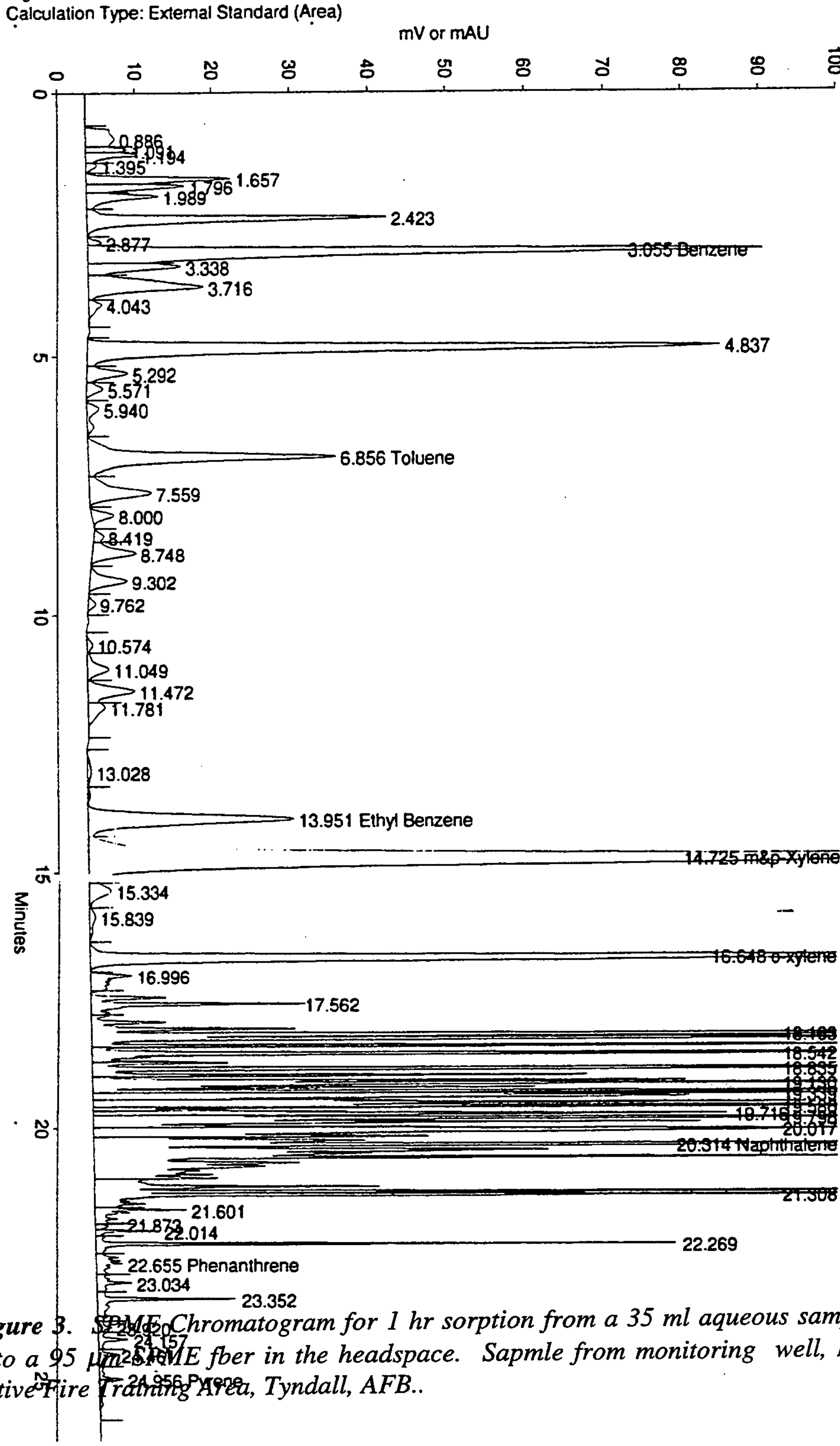


Figure 3. SPME Chromatogram for 1 hr sorption from a 35 ml aqueous sample with sorption onto a 95 μm SPME fber in the headspace. Sapmle from monitoring well, MWQ1, at FT-23 Active Fire Training Area, Tyndall, AFB..

Name: MWQ1-DL-1 hr
Type: Sample
Injection Volume: 1.0 uL

Vial: 1

Injection: 1 of 1
Injected On: 07-26-96 15:25:40

Acquisition Log
Column Pressure: N/A
Noise (microAU): 3
Run-Time Messages: None

Column Temperature (C): N/A
Drift (microAU/min): -1

Pump Flow Stability: N/A

Signal 1: Interface A
Calculation Type: External Standard (Area)

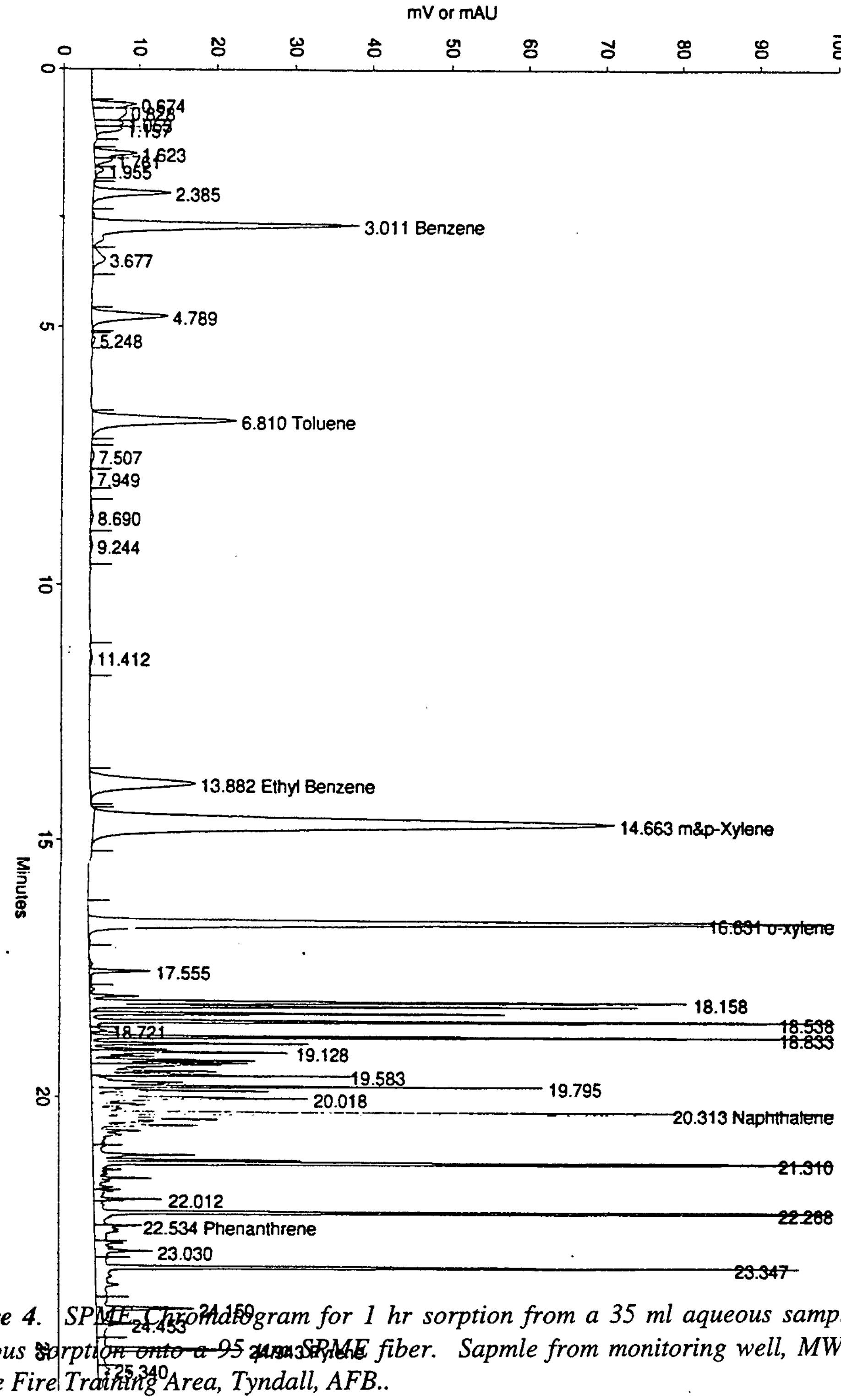


Figure 4. *SPME Chromatogram for 1 hr sorption from a 35 ml aqueous sample with direct aqueous sorption onto a 95 μm SPME fiber. Sapmle from monitoring well, MWQ1, at FT-23 Active Fire Training Area, Tyndall, AFB..*

INVESTIGATION OF NECK MODELS FOR PREDICTING HUMAN TOLERANCE TO ACCELERATIONS

Ali M. Sadegh
Professor
Department of Mechanical Engineering

The City College of
The City University of New York
140 th St. @ Convent Ave., New York, NY, 10031

Final Report for:
Summer Faculty Research Program
Escape and Impact Protection Branch
Biodynamics and Biocommunications Division
Armstrong Laboratory
Wright-Patterson Air Force Base, Ohio, 45433

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, DC

August 1996

INVESTIGATION OF NECK MODELS FOR PREDICTING HUMAN TOLERANCE TO ACCELERATIONS

Ali M. Sadegh
Professor
Department of Mechanical Engineering
The City College of The City University of New York

Abstract

During an emergency ejection from aircraft, pilots are subjected to high accelerations that may cause injuries. These injuries, especially at the neck region, are exacerbated by any additional weight that is added to the head gear of the pilot such as by night vision goggles and helmet-mounted displays. There have been several studies on head acceleration in the z and x directions all of which have investigated the rigid body dynamics of the neck.

The objective of this study is to develop a finite element model of the cervical spine that predict the stresses in each vertebra by taking into account the viscoelastic characteristics of the neck. The loads and the moments at the head point (Occipital Condyle, OC) used for the model were determined by the rigid body dynamic response of the head due to G-y and G-z accelerations as reported by Sadegh (5) and Perry (16). The experimental data used were collected from the biodynamics responses of human volunteers during a acceleration in the z and y directions on the drop tower and the sled track facility at Armstrong Laboratory at WPAFB.

Three finite elements models were developed, bulk elastic, viscoelastic and continuum. I-DEAS software were used to create the solid models, loadings and the boundary conditions. Then, ABAQUS finite element software was employed to solve the models, and thus the stresses on each vertebral level were determined.

The results indicated that the stresses in the 10G-z case were comfortable below the injury region as determined by cadaver tests. Also, the stresses in G-y accelerations increased as the magnitude of the acceleration increases. This study by no means is a complete analysis of the cervical spine and was constrained by time limitation and the scop of the study. Further studies are referred to the next report.

INVESTIGATION OF NECK MODELS FOR PREDICTING HUMAN TOLERANCE TO ACCELERATIONS

Ali M. Sadegh

INTRODUCTION

Aircraft pilots are subjected to a high acceleration environment during an emergency ejection from the cockpit. During ejection, the entire force required to accelerate the head and the helmet passes through the cervical spine. This load becomes particularly critical when additional weight is added to the helmet by night vision goggles and helmet-mounted display. Moreover, if the helmet is asymmetric, the cervical spine is subjected to additional bending and twisting. These loads can cause neck injuries and are particularly threatening to the new female pilot population. The design of helmets and helmet-mounted devices are limited by the amount of load that can be borne by the neck. Therefore, it is desirable to develop a model to predict the stresses in the cervical spine under acceleration loadings.

This model is also useful for predicting neck loads for passengers involved in automotive accidents. It has been claimed that about two-thirds of all traffic fatalities are a result of injuries to the head and neck. Whiplash is another phenomenon that can be better understood with this model. Other related areas where this model might be used include the design of helmets for motorcyclists and parachutists.

Recent advances in seat design and pilot training have added more restrain to the pilot's lumbar and thoracic regions thereby reducing the risk of lower back injuries. However, the lack of restrain in the neck and head regions has made the pilot's cervical spine vulnerable to injuries. This is due to the fact that the pilot's head must have reasonable freedom of motion in order to ensure an adequate field of view. In the ejection process, the initial orientation of the head and the direction of the acceleration vector plays an important role in the magnitude of the load that is applied to the neck. The latter statement means that the head acceleration varies considerably when the airplane is in roll or spin conditions.

There have been several studies on head and neck due to the acceleration in z and x directions, (G-z and G-x), (3 and 4). For G-y acceleration, Sadegh (5) developed and validated a head and neck model. He employed the experimental data collected from the biodynamics responses of human volunteers during an acceleration in the y direction on a sled at the sled track facility at Armstrong Laboratory at WPAFB. He employed the Articulated Total Body (ATB) software and determined the loads and torques at the neck/head and the neck/torso joints. In all these studies the global rigid body dynamic response of the head and the neck due to the acceleration was determined. In addition the global loads and torques at the head point (OC) and the neck points were also determined. However, the local loads (stresses) on each vertebra were not calculated - in any of these studies.

There are five major mechanisms involved in cervical injuries; compression, flexion, extension, rotation, and lateral flexion. However, upper cervical spine (C1 and C2) injuries are mainly due to hyperextension, and their dislocations are fatal. The intervertebral discs, joints and ligaments are very resistant to compression, distraction, flexion and extension, but very vulnerable to rotation and horizontal shearing forces such as in G-y accelerations, Roaf, (10). According to Roaf (10), the clinical approach of a cervical dislocation or fracture-dislocation, usually attributed to hyperflexion, is really the result of rotation. Belyschko et. al. (11) reported that the maximum voluntary static neck reaction is about 1.13×10^8 dynes in tension and 1.11×10^8 dynes in compression. A limited amount of strength data for individual components of the neck can be found in the literature.

The purpose of this report is to develop finite element (local) models for the cervical spine that include seven cervical spines, the six intervertebral discs as well as the ligaments. Three finite elements models were developed, bulk elastic, viscoelastic and continuum. I-DEAS software were used to create the solid models, loading and boundary conditions. Then, ABAQUS finite element software was employed to solve the models, and determine the stresses on each vertebral level.

OBJECTIVES:

The aim of this study was to develop a three dimensional finite element models of the cervical spine to predict the local loads due to the G-y and G-z accelerations which could not be provided by the global rigid body dynamic model that were developed in previous studies. Specifically, the focus of this study was :

1. to develop a layered elastic and viscoelastic bulk models of the neck and determine the stress and displacement at each layer, and
2. to develop a continuum model of the cervical spine with ligament attachments that will represent a more realistic model of the neck.

The finite element models developed in this study are capable of simulating the response of the muscle-skeletal structures of the neck when it is subjected to the forces due to G-z or G-y acceleration in an ejection process. The first and second models were intended to be a bulk representation of the neck and the third model to have more anatomical detail of the cervical spine. The stresses determined from these models are compared to the human tolerance level . Analysis of these models is the first step towards understanding of the local stresses in each vertebra.

CERVICAL SPINE

A typical cervical vertebra is shown in Figure 6. The characteristic feature of the cervical vertebrae is a foramen in the transverse processes for the passage of the vertebral artery, vein and sympathetic nerves. The first cervical vertebra, known as the atlas, supports the skull. It has no vertebral body and no spinous process, but is made up of two lateral masses and two arches. The second vertebra, called the axis or epistropheus, has a tooth like process, conical in shape. The third to sixth vertebrae have the typical, standard shape shown in Figure 6. They have small vertebral bodies that are broader from side to side than they are from front to back. The seventh vertebra has a long, nearly horizontal spinous process which serves as an attachment point for many neck muscles. The ligaments of the cervical spine bind the vertebrae together as they do in the rest of the spine, and together with the paracevical muscles prevent

any motion that would injure the spinal cord and nerve roots. Most of the axial rotation of the head on the neck occurs between the first two vertebra, the atlas and the axis.

Extension in the cervical spine is limited at the upper end by the superior facets of the atlas whose posterior edges lock into the occipital condylar fossae. Flexion is stopped just after the cervical convexity is straightened; the limiting factor is the contacting of the overhanging lops of the bodies of the vertebrae with the wall of the subjacent vertebral bodies. In this study our interest is in the lateral flexion. The lateral flexion in the lower cervical spine (C2-C7) is always coupled with a certain amount of axial rotation. This coupling is such that during lateral bending to the left the spinous processes go to the right, and during lateral bending to the right they go to the left.

EXPERIMENTAL DATA

Sadegh (5) employed experimental data that were collected from the biodynamic responses of human volunteers during an acceleration in the z and y directions at the drop tower and the sled track facility located at the Escaped and Impact Protection Branch at Armstrong Laboratory at WPAFB. The sled test facility employs an Impulse Accelerator (Shaffer 1976) that consists of a gas powered actuator which accelerates a sled on a two-rail track. The volunteers were placed in a chair that is mounted on the sled facing perpendicular to the direction of the track. Two sets of three orthogonal linear accelerometers were located in a chest pack and a mouth pack. These accelerometers collected the x, y and z accelerations of the torso and head as a function of time during the acceleration impulse. For the G-y acceleration, the sled was subjected to the acceleration pulse of a half-sine with peak acceleration ranging from 4G to 7G and duration ranging from 31 ms to 250 ms. For the G-z acceleration, the subject on the drop tower was subjected to an acceleration pulse of +10G-z.

Because of safety rules the test subjects were not subjected to an acceleration of more than 10G. However, cadavers were used to estimate the maximum tolerance and the injury region of the cervical spine. Buhrman and Perry (15) reported that for the neck in flexion based on the +Gx direction, the maximum responses of a cadaver without producing ligament or bone

damage at the occipital condyle were estimated as: M-y of approximately 1700 in-lb, shear load Fx of 450 lb and compression load Fz of 400 lb.

Although a great number of cases for G-z and G-y were available, due to the lack of time and large volume of data to be processed, only one case for the G-z acceleration was considered. A complete analysis will be presented in the next report.

To determine the forces and the moments at the head point (OC) for the G-z and G-y accelerations a head/neck model consisting of three segments, namely, Head, Neck, and Upper Torso was developed and reported in Sadegh (5) and Perry (16). For these models, the weight and physical and geometric information of each segment were taken from GEBOD software. The GEBOD program is an interactive computer program that produces the human and dummy body description data used by the ATB model. In this model the head segment is joined to the neck by "Head Pin" (HP) joint and the neck is connected to the upper torso by the "Neck Pin" (NP) joint.

In this analysis, eight cases were considered. The first two cases were studied using the bulk elastic and viscoelastic models. The third model was used for the remaining six cases that involved detailed geometric descriptions of the vertebrae and the discs. The calculated loads and moments used in this analysis are shown in Table 1.

THE MODELING

Three models were developed. In the first model, the cervical spine including the soft tissue and the muscles were considered as a bulk column of seven vertebra and six discs. However, in order to apply the loading and the boundary conditions to the model and minimize the stress concentration, one extra disc was added to both the top and the bottom of the model. The top disc was a thin rigid material that was included in the model for better distribution of the applied loads and the torques. The bottom thick disc represented the T1 vertebra. The cross section of the body of a vertebra is an ellipse that is very close to a circle, with an average radius of 9 mm, as given in Williams and Belytschko [9]. However, to compensate for the arch and processes of the vertebrae and the soft tissues the radius of the column was assumed to be

15.8 mm which is larger than the radius of the body of a vertebra. In this model 589 quadratic ten-noded tetrahedron elements and 1568 nodes were used. The loads were applied to a rigid plate on top of the C1 in order to avoid the stress concentration. The model was constrained at T1 at the inferior of the seventh cervical vertebra. The material properties of bone used in this model are given in Table 2.

In the second model the material properties of the discs were changed to a viscoelastic material defined by Prony series representation of the normalized shear and bulk relaxation moduli, see ABAQUS [8]. The properties of the muscles and ligaments are given in Williams and Belytschko (9) and Woo et. al. (14). The viscoelastic coefficients of the materials are given in Table 2.

The third model was a detailed anatomical description of the cervical spine as shown in Figure 6. The complete cervical vertebrae in sagittal and posterior views are shown in Figure 7 and 8, respectively. These models were generated using I-DEAS software from SDRC Inc., running on a DEC workstation. I-DEAS is a large solid modeler, with pre and postprocessor software that interacts with ABAQUS software. ABAQUS software is a research oriented nonlinear finite element solver with wide variety of linear and nonlinear elements for many different kind of engineering analysis. Each vertebra and disc was carefully created using the dimensions given in the Cervical Spine Research Society Book [17]. Also, the relative orientation of the vertebrae and the discs were carefully generated to resemble the anatomical orientation as given in [17]. Initially, close to 3000 quadratic ten-node tetrahedron elements and over 5000 nodes were used. When the model was processed, however, it was found that ABAQUS solver took a rather long time (4 hours) to complete the calculation, probably due to the nonlinear elements (such as viscoelastic beam elements). A simpler model with linear tetrahedral elements was then developed. The new model employed 973 four-noded tetrahedral elements with 2431 nodes.

In order to simulate the ligaments 36 truss elements with ligament material properties were used to connect the inferior section of the posterior arch of each vertebra to the superior of the posterior arch of the adjacent vertebra. By selecting the right material properties for

Table 1. Description of six cases

unit			lb			in-lb	
item	cell	accel.	FX	FY	FZ	MX	MY
First:bulk elastic model					133		309
Second: bulk viscoelastic model					Fig. A		Fig. A
Third model:							
case 1	cadaver	maximum load	450	-400		1700	
case 2	CB#2368	10 Gz	25		-200	320	
case 3	A #4128	4Gy	20	-55		-120	-80
case 4	B #4147	5Gy	30	-100		-200	-90
case 5	C #4165	6Gy	40	-110		-230	-90
case 6	D #4185	g	55	-130		-310	-120

Note: Figure A is in the Figure 15 of the report Sadegh (5)

Table 2. The material properties

Elastic:	Vertebra	Disc	Ligament
Modulus of Elasticity	200 MPa	4.2 MPa	2.2 MPa
Pisson's ratio	0.3	4.5	4.5

Viscoelastic properties

Shear relaxation	Bulk Relaxation	Relaxation time
g1=0.3991,	k1=0.7, and	t=3.4519
g2=0.3605,	k2=0.149, and	t=2000
g3=0.1082,	k3=0.150 and	t3=7000

these truss elements we simulate the ligaments around the bodies and the posterior arch of the vertebrae.

As the boundary conditions for this model the nodes on the inferior surface of the body of the C7 were constrained. The loads were distributed over the superior surface of the C1. The loads shown in Table 1 were applied to different models. Because of the applied loads and constraints on the C1 and C7 the results of these two vertebrae should be ignored.

THE RESULTS

The elements, nodes, boundary conditions and the loads of different models created in I-DEAS were used to create the input files for ABAQUS finite element method program. The results of each run, i.e. the stresses and displacements of each layer representing the discs and the vertebrae, were tabulated and analyzed. Due to the large number of tables and graphs generated and the page limitation for this report only a representative selection numbers of graphs and figures are presented here.

The maximum tensile stress and compressive stresses on the vertebrae of the first bulk elastic model are shown in Figures 1 and 2, respectively. As indicated in the modeling section, the first and the last points on these figures are not part of the cervical spine and should be ignored. The overall deflection of the model due to the load is shown in Figure 3. The maximum tensile stress and compressive stresses on the vertebrae of the second viscoelastic bulk model are shown in Figures 4 and 5, respectively. These curves are similar to that of Figures 1 and 2, except the magnitude of the stresses are lower for the viscoelastic model. This is due to the fact that the shock effect of the load is absorbed by the viscoelastic properties of the discs. Clearly, as the relaxation time tends to go to infinity the stresses predicted by the two models approaches equality.

The oblique view of the third model given by solid modeling is shown in Figure 6. The sagittal view and the posterior views of the finite elements of the model are shown in Figures 7 and 8, respectively. The truss elements representing the ligaments are shown in Figure 7 as connecting lines between the two nodes of the posterior arches of adjacent vertebrae. The

results of maximum principal stress S33 on the discs and the vertebrae of the third model for the cases 1 through 6 are shown in Figures 9 to 11. The results are presented for the anterior and posterior sections of the discs and the vertebral bodies since these two sections are vulnerable to fracture and damage. Note that, the first two cases of the third model were subjected to G-z acceleration and cases 3 to 6 of the model were subjected to the G-y acceleration. Therefore, the curves for cases 1 and 2 are separated from the rest of the cases. The magnitude of the stresses for the first case is high because the cadaver was loaded to measure the maximum tolerance and the injury level. Stresses for this case are therefore considered as the injury tolerance and failure level.

The maximum principal stresses S33 for the posterior section of the intervertebral discs are shown in Figures 9a and 9b. Discs are labeled 1 through 6 representing the intervertebral of vertebrae C1 to C7, respectively. Similar stresses for the anterior section of the discs are shown in Figure 9c and 9d. The maximum principal stresses S33 for the posterior and interior sections of the vertebral bodies are shown in Figures 10a through 10d. The stresses in the anterior and posterior sections of the vertebral bodies have some fluctuation because of the curvature, geometry and ligament attachment of the cervical spine model. The overall trend, however, is toward greater tension in the posterior section and more compression in the anterior section. This is an expected since the cervical spine is a complex curved cantilever beam in the global model.

To understand the effect of the increased in acceleration on the stress levels in the vertebral bodies and the discs, these variations are plotted in Figures 11a to 11d. These plots are for G-y acceleration since only two point data for the G-z accelerations were available and complete data for G-z were not analyzed. Figures 11a and 11b depict the maximum principal stress variation of the posterior and anterior sections of each disc as a function of G-y accelerations. Likewise, for the vertebral body the maximum stresses are shown in Figures 11c and 11d. These figures indicate that in the posterior sections of the discs and the vertebral bodies the increase in the G-y acceleration causes the most stress variation on discs 1 and 6 and the vertebra C1 and C7.

CONCLUSION

The first bulk elastic model was a gross model of the neck that did not yield accurate results. The enlarged radius of 15.8mm was a geometrical average of the vertebrae that includes the arches and the processes. While the stress levels of the model and the stress variation seems reasonable the justification for the enlarged diameter of the model is difficult. This is due to the fact that actual compensation for the soft tissue and the muscles attachment was not analyzed. However, the model could be used as a quick and simple model to find the critical accelerations and stresses.

The second model, the viscoelastic bulk model, produced less stresses in the discs and the vertebral bodies. This is due to the fact that the damping effect of the discs materials absorbs some of the impulse loads. If the time is extended to infinity the final loads, after relaxation, will approach the level of the first model.

Geometrically, the third model is an accurate model of the cervical spine. While the effect of the soft tissues and the ligaments were compensated by the truss elements in the model more detailed nonlinear spring and dashpot elements could be used to simulate the neck more realistically. Comparison of cases 1 and 2 of the third model reveals that the 10G acceleration are comfortably below the maximum stresses of the cadaver. As the G-y acceleration increases in cases 3 to 6 the maximum stresses in the discs generally increases. However, the stresses in the vertebrae fluctuates. This is due to the fact that the geometry of the cervical spine is complex and nonlinear. It resembles a curved beam that is subjected to axial and bending load. Since the points of application of the loads are not at the centroid of the cross-section the stresses found in the vertebral bodies were not uniform.

This study by no means is a complete analysis of the cervical spine. In fact the third model should be further modified to include spring elements. Additional studies are needed to compare these maximum stresses to the soft tissue damage since the stresses are substantially below the cadaver stresses. Further, the issue of the dynamic response of the neck to the impulse load should be addressed. These analyses were beyond the time limitation and scop of this study and are deferred to the next report.

REFERENCES

1. Guill, F.C. and Herd, G.R., "Aircrew Neck Injuries: A new, or an existing misunderstood phenomenon?", Neck Injury in Advanced Military Aircraft Environment, AGARD, conference Proceedings No. 471, pp,9.1-12, 1990.

2. Knox, F.S., Buhrman, J.R., Perry, C.E., Kaleps, I."Interim head/neck criterion", Armstrong Laboratory Consultation report, WPAFB, OH, Dec., 1991.

3. Miller H.L. and Schneck D.J.,"Biodynamic responses to emergency ejections under conditions of added head mass", WPAFB, research contract, 1993.

4. Perry, C.E., "Vertical impact testing of two helmet-mounted night vision systems", AL/CF-SR, 1994-0013

5. Sadegh, A. " analysis of loads on the neck and head joints due to G-y acceleration" SFRP report, AL/CFBE, WPAFB, 1995.

6. Foust D.R., Chaffin, D.B., Snyder, R.G. and Baum, J.K., "Cervical range of motion and dynamic response and strength of cervical muscles," Proceedings of 17th Stapp car conference, paper #730975, 285-309, 1973.

7. Johnson , RW., Crelin, ES, White III, A.A., Panjabi, M.M. and Southwick, W.O., "Some new observations on the functional anatomy of the lower cervical spine", Clin. Orthop: 3, 192-200, 1975.

8. ABAQUS user and theoretical manuals, HKS Inc, 1995.

9. Williams J. and Belytschko T. "A Dynamic model of the cervical spine and head", AFAMRL-TR- 81-5.

10. Roaf R. "A Study of the mechanics of spinal injuries", J. Bone Joint Surg., 42B, 810-823, 1960.

11. Belytschko T. Rencis M. and Williams, "Head-Spine structure modeling: enhancements to secondary loading path model and validation of head-cervical spine model" AAMRL-TR-85-019, 1985.

12. Visarius H.,"On the experimental and mathematical modeling of the viscoelastic behavior of the human spine", Dissertation, Wayne State Univ., 1994.

13. Farfan H., Gracovetsky, S., and Helleur C." Cervical spine analysis for ejection injury prediction", US AFOSR #81-0012, 1982.

14. Woo SLY, Johnson GA., Smith B." Current concepts in modeling of ligaments and tendons, ASME Bioeng. Conf. 1993, BED-Vol 24, 298, 1993.

15. Buhrman, J. and Perry, C. "Human and Manikin Head/Neck Response to +Gz Acceleration when encumbered by helmets of various weights", Aviation, Space, and Environmental Medicine, P 1086-1090, December, 1994.

16. Perry, C. "Gz Impact acceleration test results for human subjects and manikin", internal communications of the test results., 1996.

17. The Cervical Spine", The Cervical Spine Research Society, Second edition, Lippincott Company, 1995.

18. White A. and Panjabi M. "Clinical biomechanics of the Spine", Lippincott Company, second edition, 1995.

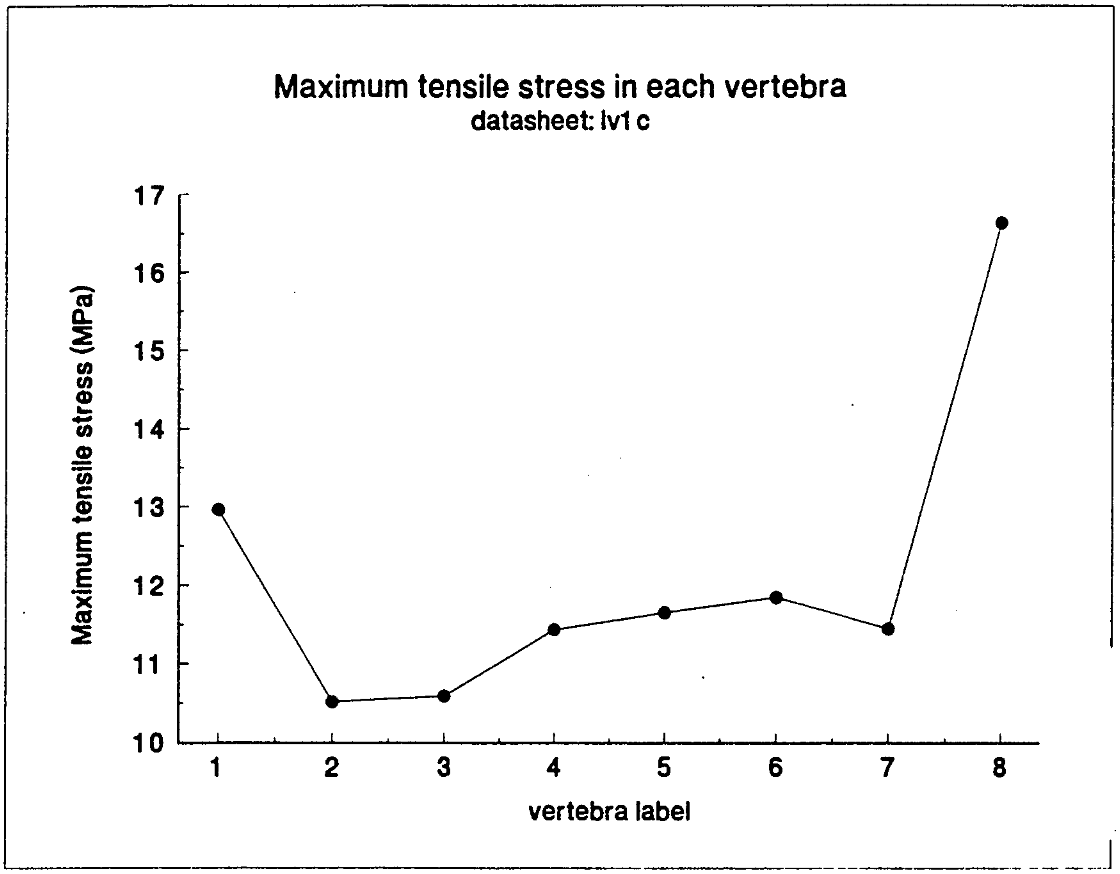


Figure 1: Bulk elastic model, tensile stress vs. vertebra

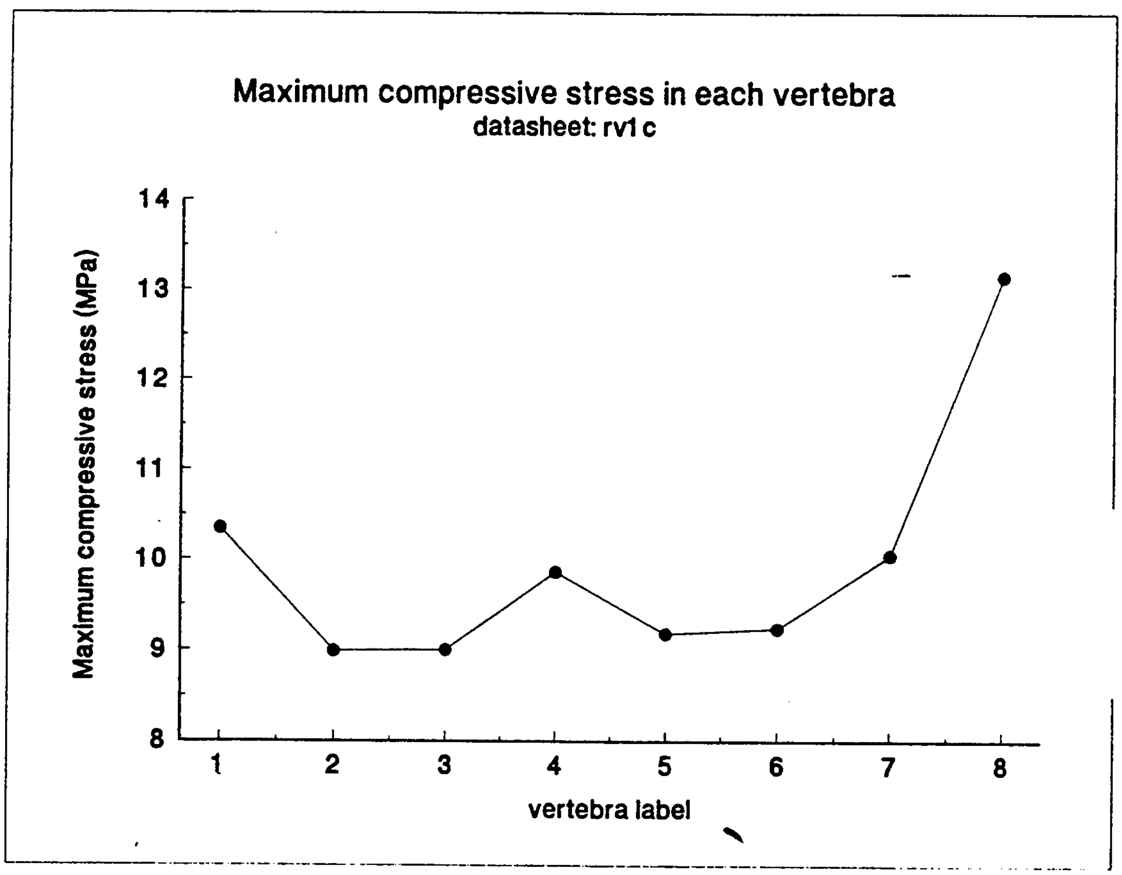


Figure 2: Bulk elastic model, compressive stress vs. vertebra

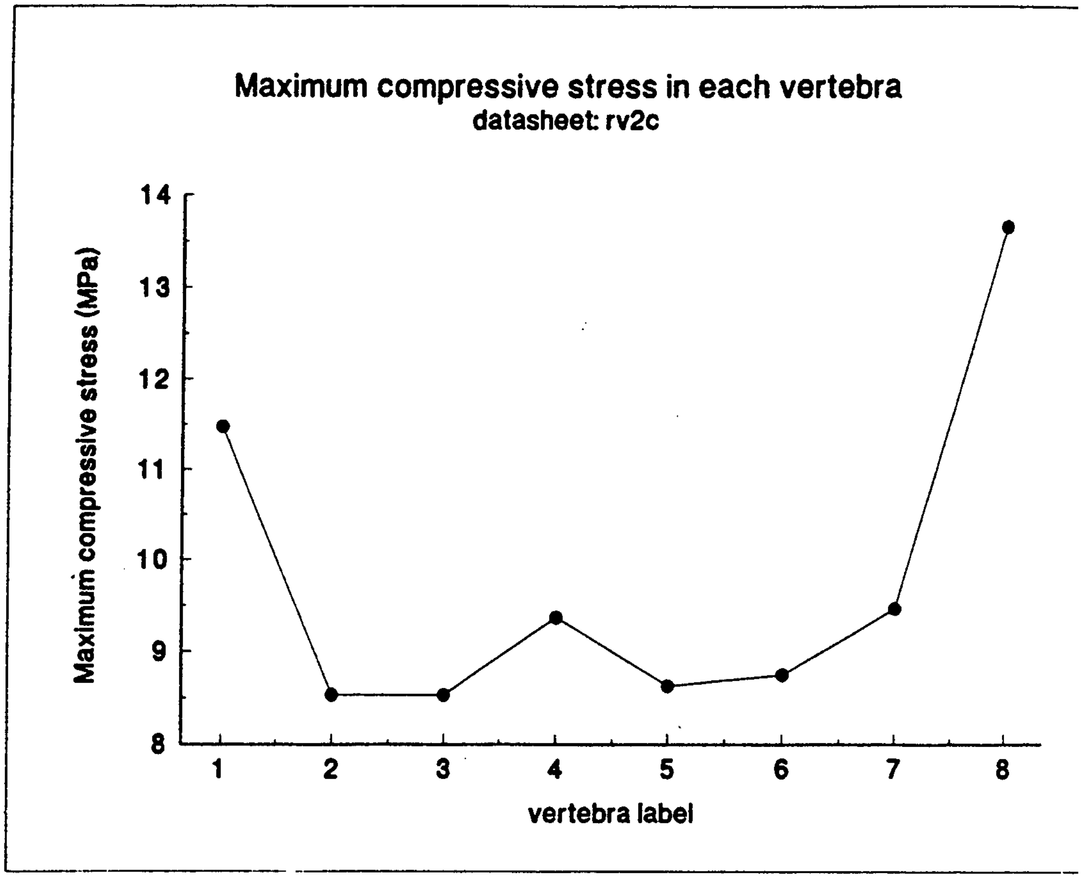


Figure 5: Bulk viscoelastic model, compressive stress vs. vertebra

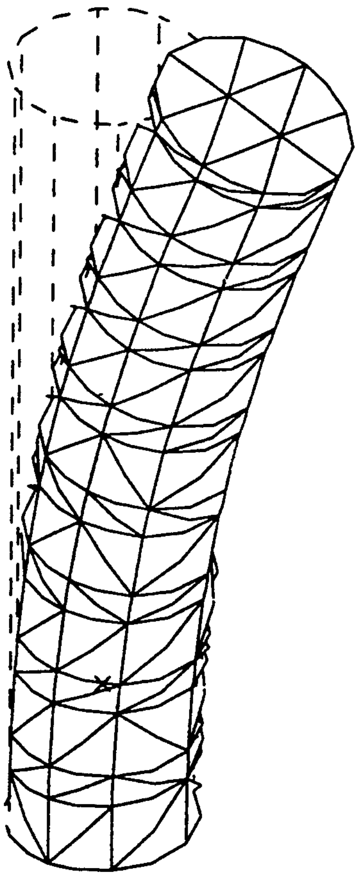


Figure 3: Bulk elastic model, Global displacement

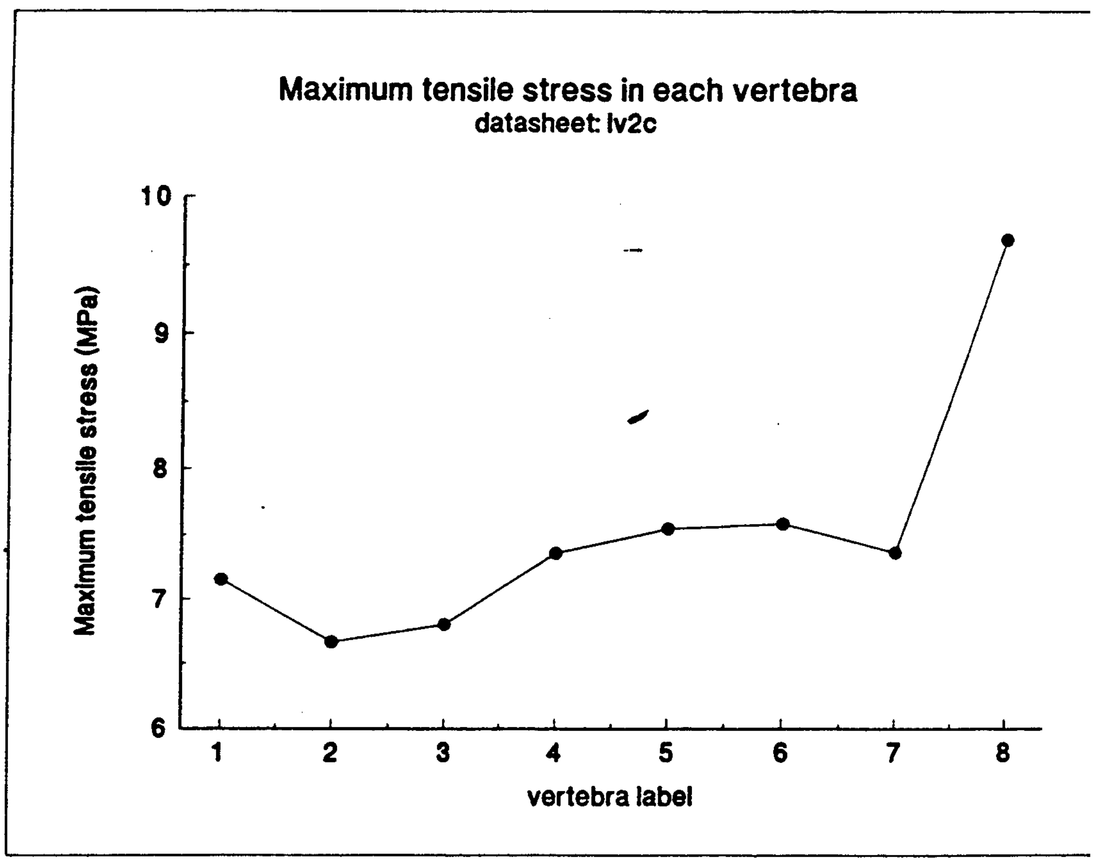


Figure 4: Bulk viscoelastic model, tensile stress vs. vertebra

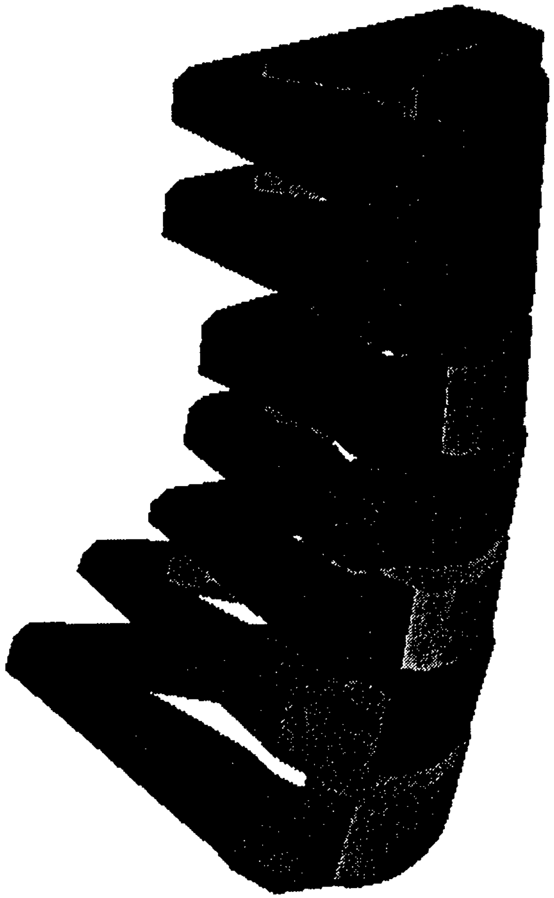


Figure 6: Oblique view of the third model

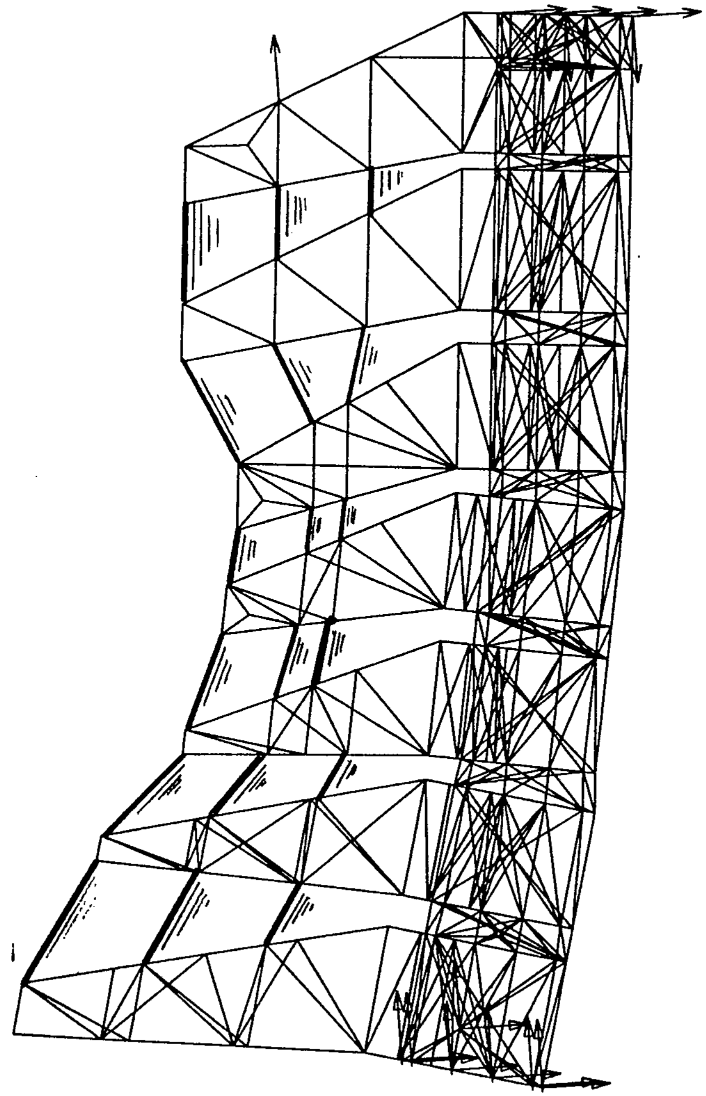


Figure 7: Sagittal View of the third model

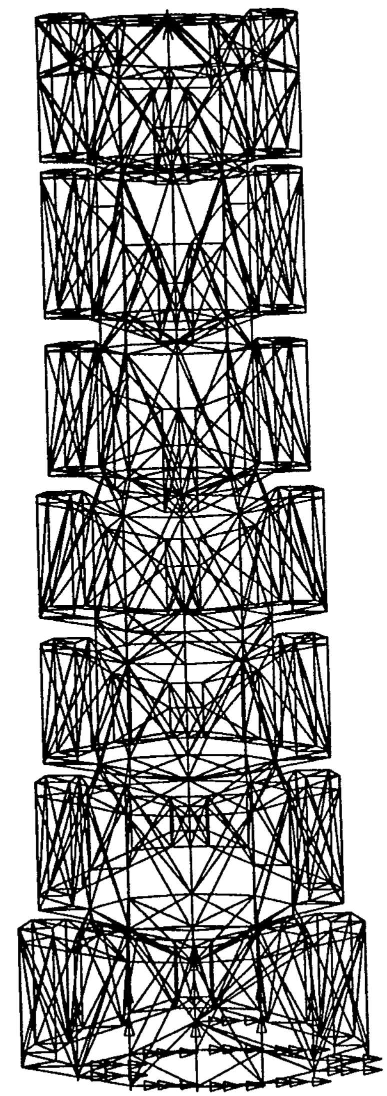


Figure 8: Posterior view of the third model

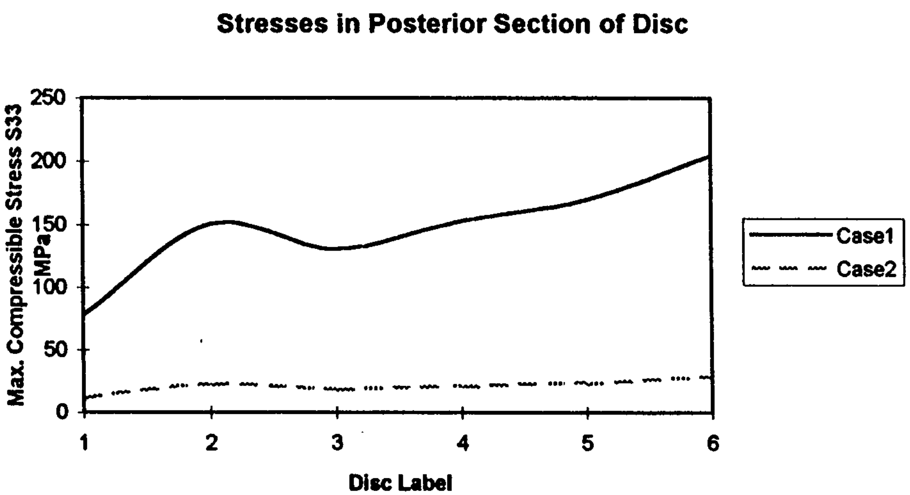


Figure 9-a

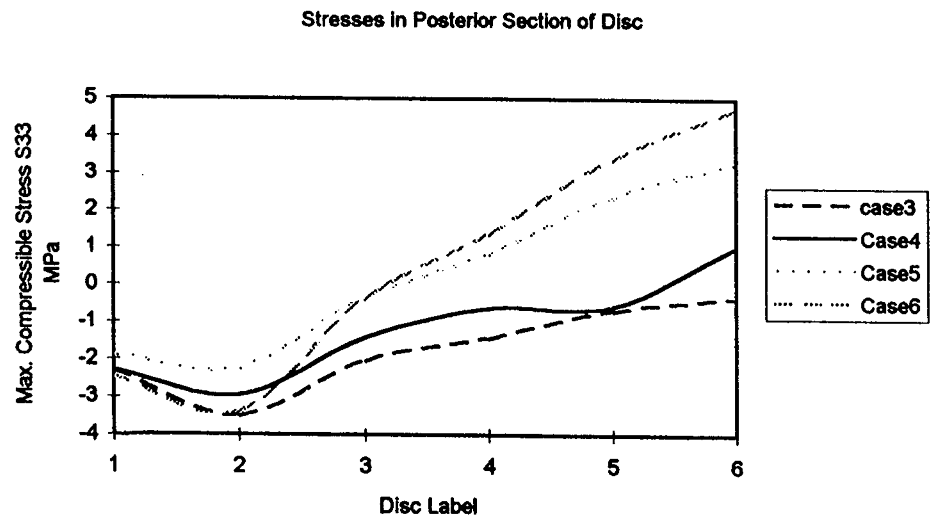


Figure 9-b

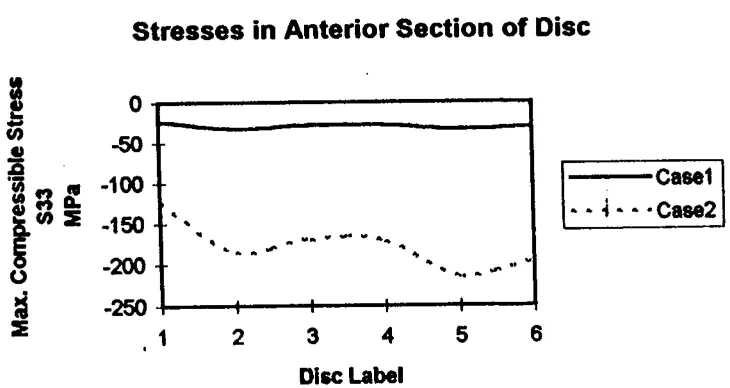


Figure 9-c

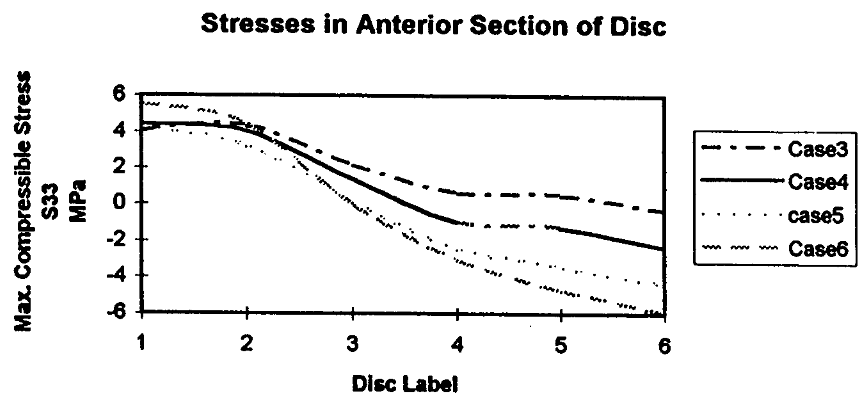


Figure 9-d

Figure 9: Third model: Maximum principal stresses in anterior or posterior of discs

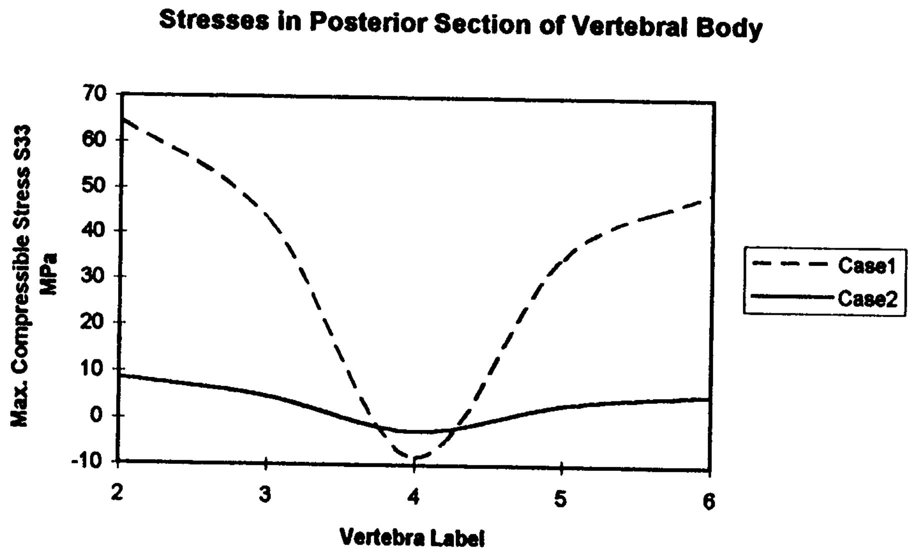


Figure 10-a

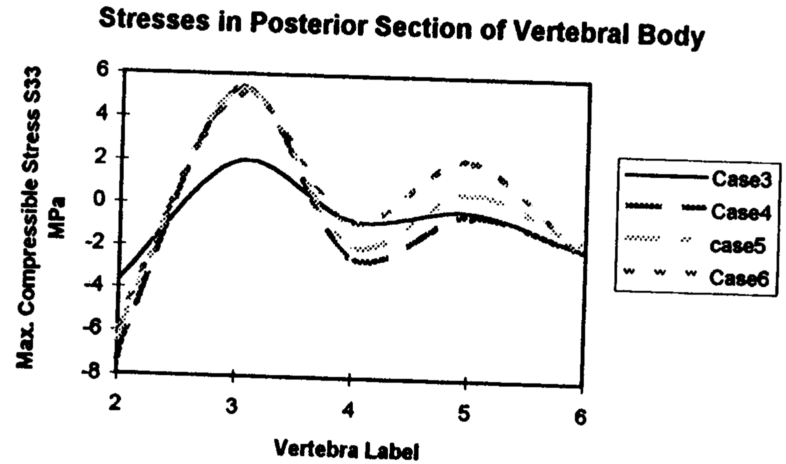


Figure 10-b

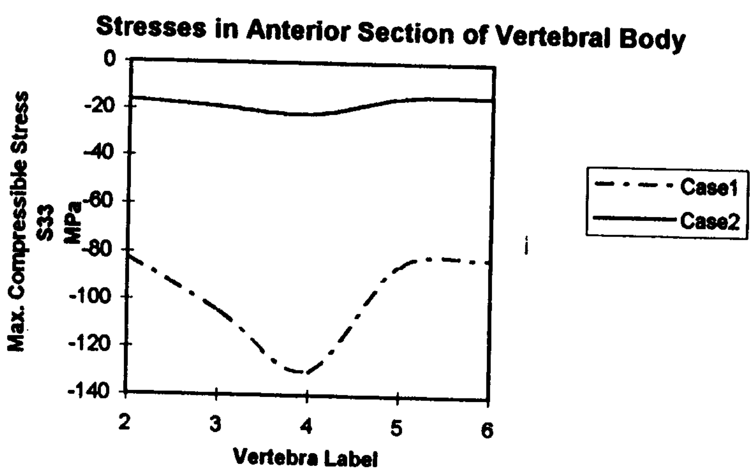


Figure 10-c

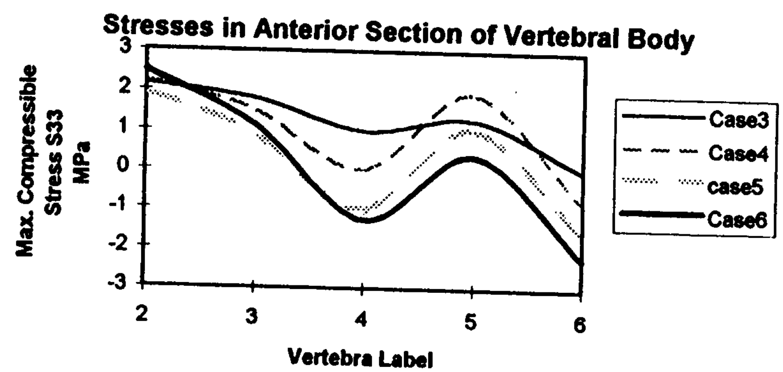


Figure 10-d

Figure 10: Third model: Maximum principal stresses in anterior or posterior of vertebra

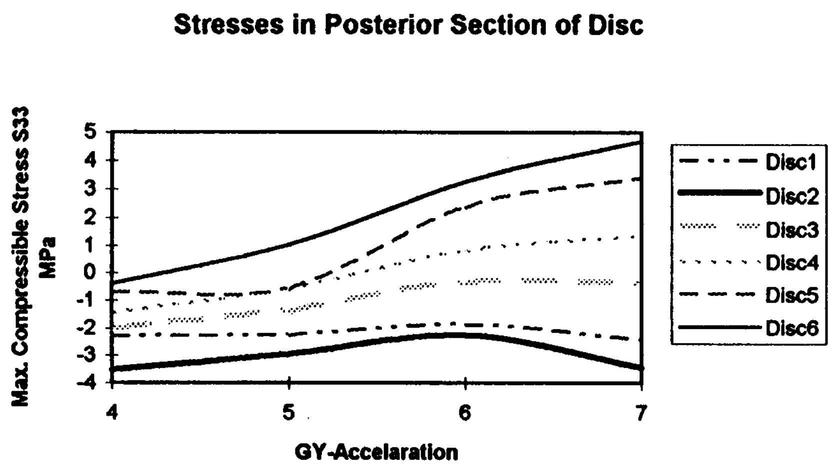


Figure 11-a

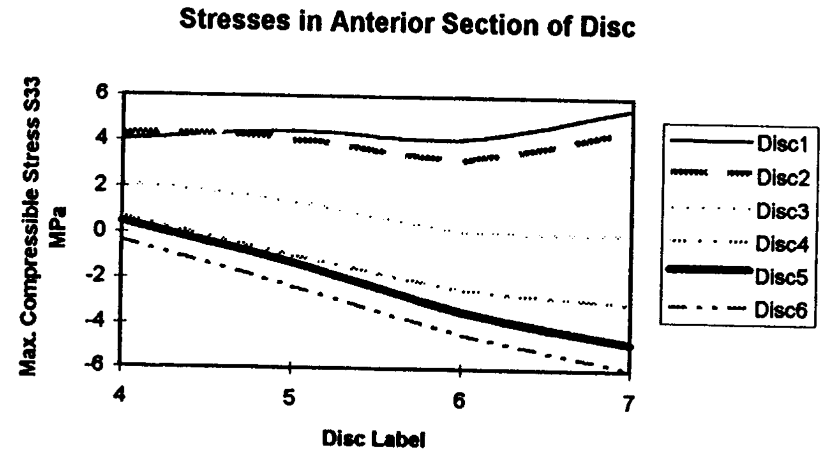


Figure 11-b

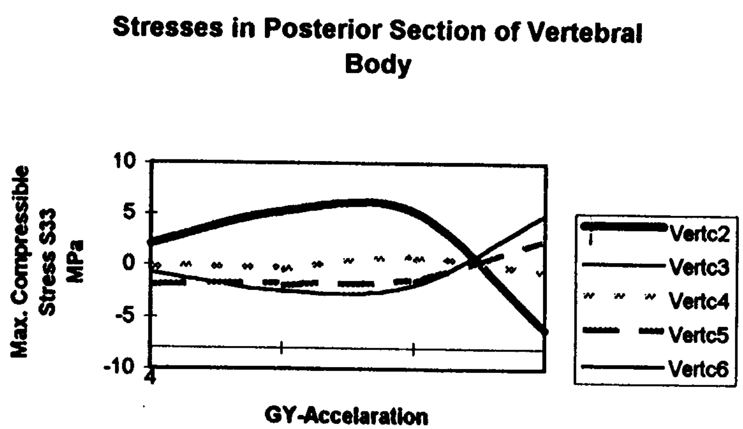


Figure 11-c

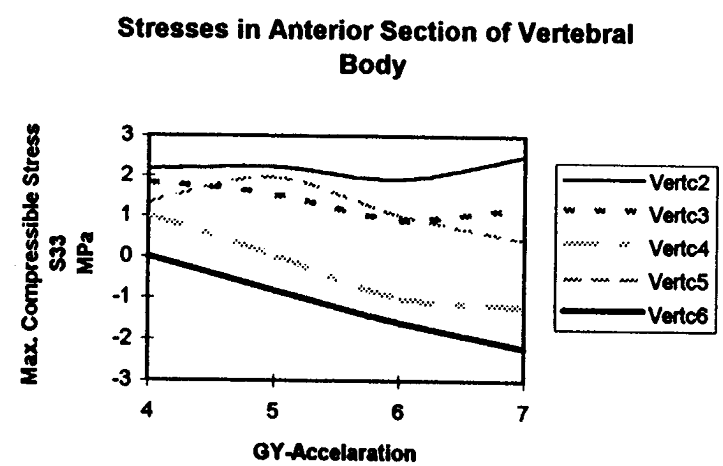


Figure 11-d

Figure 11: Third model: Maximum principal stresses in anterior or posterior of discs or a vertebra v.s. the G-y acceleration

TRUNCATED BIVARIATE EXPONENTIAL MODELS

Kandasamy Selvavel
Associate Professor
Department of Mathematics
and
Computer Science

Claflin College
700 College Avenue
Orangeburg, SC 29115

Final Report for:
Summer Faculty Research Program
Armstrong Laboratory

Sponsored by:

Air force Office of Scientific Research
Bolling Air Force Base, DC

and

Armstrong Laboratory
Brooks Air Force Base
San Antonio, TX

September 1996

TRUNCATED BIVARIATE EXPONENTIAL MODELS

Kandasamy Selvavel
Associate Professor
Department of Mathematics and Computer Science
Claflin College

Abstract

Truncated distributions have been the subject of study for many years. But, most of the work done so far involves univariate truncated families. In this study, we consider two different forms of truncated bivariate distributions. Under certain conditions it is shown that one of these distributions has exponential marginal. Method of moments and uniform minimum variance estimators are used to estimate the unknown parameters. We give an estimator of the probability that Y is less than X. Moreover, some graphs are given to compare the fit of the exponential model.

TRUNCATED BIVARIATE EXPONENTIAL MODELS

Kandasamy Selvavel

1. INTRODUCTION

Estimation of Parameters of Truncated Bivariate Exponential Models

Exponential distributions play main role in life testing, reliability and other important field of studies. In the bivariate case, several authors have considered the problem of estimation of parameters for regular distributions. Estimation of probability of Y less than X when X and Y are independent exponential variates has been considered by many authors. Tong (1974, 1975) derived two expressions for the uniform minimum variance unbiased estimator (UMVUE) of the probability that Y less than X for the negative exponential and gamma distributions in closed forms, and Beg (1980) considered two-parameter exponential distributions. A correction was made by Johnson (1975) in one of the expressions for UMVUE given by Tong. Kelly and Schucany (1976) derived the maximum likelihood estimator and the UMVU estimator for the probability that Y less than X. Harris (1967) considered some reliability applications of bivariate exponential distributions. Arnold and Strauss (1988) considered method of moment estimators for bivariate distributions with exponential conditional.

Almost all the work done so far involves regular multivariate exponential distributions. But, very few results are available in truncated case. The distributions for which one or both of the extremities of the range of the distribution are functions of the unknown parameters are called truncated distributions. In our study, we consider two different left truncated bivariate exponential distributions. Under certain conditions we show that one of these truncated distributions has exponential marginal. Method of moments and uniform minimum variance unbiased estimators are used to estimate the unknown parameters.

Further more, we illustrate our results using a published data from a study of 2, 3, 7, 8-tetrachlorodibenzo-p-dioxin exposure among 46 members of Operation Ranch hand, the Air Force unit responsible for the aerial spraying of Agent Orange in Vietnam. Some graphs are given to compare the fit of the exponential model.

2. UNIFORM MINIMUM VARIANCE UNBIASED ESTIMATION OF PARAMETERS OF BIVARIATE EXPONENTIAL DISTRIBUTIONS

Several authors have studied the problem of estimation of parameters of bivariate exponential distributions when X and Y are independent variates. In this study, we estimate the parameters when X and Y are dependent variates. It is well known that, if the extremities of the range of a probability density function depends on unknown parameters, then the classical regularity conditions for maximum likelihood estimators (mles) are not satisfied. Hence, we derive uniform minimum variance unbiased estimators of parameters of truncated bivariate exponential distribution for this case. More specifically, we restrict our attention to the truncated bivariate probability density function (pdf) of the form

$$f_1(x,y)=q(\theta_1,\theta_2)h(x,y), \qquad \begin{array}{l}\theta_1<x,\ \theta_2<y,\\ \theta_1>\theta_2,\end{array} \tag{2.1}$$

where

$$q(\theta_1,\theta_2)=\frac{6e^{3\theta_1}}{3e^{3\theta_1-\theta_2}-1}$$

and

$$h(x,y)=e^{-x-y-\max(x,y)}.$$

Let (x_i,y_i), $i=1,2,\ldots,n$ be a random sample from the probability density function f_1. Also, let $x_{1:n}<x_{2:n}<\ldots<x_{n:n}$ and $y_{1:n}<y_{2:n}<\ldots<y_{n:n}$ be the corresponding order statistics of x_i's and y_i's respectively.

Taking $\alpha=-1$, $\beta=0$, $\gamma=0$ and $\delta=1$ in Theorem 1 of Selvavel (1992), we have the following lemma.

Lemma1: Let (x_i,y_i), $i=1,2,\ldots,n$ be a random sample from probability density function f_1. Then the uniform minimum variance unbiased estimator (UMVUE) of an unbiased estimable function $g(\theta_1,\theta_2)$ is given by

$$\phi(x_{1:n},y_{1:n})=\frac{q^{n+2}(x_{1:n},y_{1:n})}{n\{(n1)q_1(x_{1:n},y_{1:n})q_2(x_{1:n},y_{1:n})+q^3(x_{1:n},y_{1:n})h(x_{1:n},y_{1:n})\}}\cdot\frac{\partial^2}{\partial x_{1:n}\partial y_{1:n}}\left(\frac{g(x_{1:n},y_{1:n})}{q^n(x_{1:n},y_{1:n})}\right),$$

where

$$q_1(x_{1:n},y_{1:n})=\frac{\partial q(x_{1:n},y_{1:n})}{\partial x_{1:n}}$$

and

$$q_2(x_{1:n},y_{1:n})=\frac{\partial q(x_{1:n},y_{1:n})}{\partial y_{1:n}}.$$

Now, using Lemma 1 to pdf (2.1), the UMVU estimators of θ_1, θ_2 and $q(\theta_1,\theta_2)$ are given by

$$\hat{\theta}_1=\frac{3(n-1)x_{1:n}e^{x_{1:n}}(3-e^{y_{1:n}-1})-3e^{x_{1:n}-y_{1:n}}-3(n-1)x_{1:n}+1}{6(3-2n)e^{x_{1:n}-y_{1:n}}+3n-5},$$

$$\hat{\theta}_2=\frac{y_{1:n}(1-3n+6ne^{x_{1:n}-y_{1:n}})+(3-e^{y_{1:n}-x_{1:n}})(1-2e^{x_{1:n}-y_{1:n}})}{6(3-2n)e^{x_{1:n}-y_{1:n}}+3n-5},$$

and

$$q\widehat{(\theta_1,\theta_2)}=\frac{6(n-1)}{n}\frac{[3ne^{3x_{1:n}}(e^{x_{1:n}-y_{1:n}-1})+2e^{3y_{1:n}}\{3e^{x_{1:n}}-2(\bar{3}e^{x_{1:n}-y_{1:n}-1})\}]}{[6(2n-3)e^{x_{1:n}-y_{1:n}}-3n+5]}.$$

It is possible to find the UMVU estimator of the probability that Y is less than X in a closed form. Since this expression is very complicated, it is not included in this study.

Michalek (et al 1996) studied the reliability of the serum dioxin measurements using paired serum dioxin measurements in 46 enlisted Ranch hand veterans participating in the Air Force Health Study. We now fit our exponential model to this data set, and estimate the parameters using the

above expressions. Using table I with $X_{1:n}$=3.3, $Y_{1:n}$=3.27 and n=46, we obtained

$$\hat{\theta}_1 = 26.72$$

$$\hat{\theta}_2 = -1.50$$

and

$$q(\hat{\theta}_1, \hat{\theta}_2) = 3954.11.$$

These estimators of the left truncation points suggest that this model doesn't fit very well for the reliability data set. This model is mainly applicable in system reliability studies.

In the next section, we use a modified version of Marshall and Olkin (1966) bivariate model.

3. A MODIFIED VERSION OF MARSHALL AND OLKIN BIVARIATE MODEL

A model based on the exponential distribution has been used by Marshall and Olkin (1966) to determine a bivariate distribution. But, this model does not have exponential marginals. In our study, we use a modified truncated version of Marshall and Olkin model. Under certain conditions we show that this truncated distribution has an exponential marginal.

More specifically, we consider the truncated bivariate probability density function (pdf) of the form

$$f_2(x,y) = \frac{\lambda^2 e^{-\lambda(x+y)}[1+\alpha(2e^{-\lambda x}-1)(2e^{-\lambda y}-1)]}{e^{-\lambda(\theta_1+\theta_2)}[1+\alpha(1-e^{-\lambda\theta_1})(1-e^{-\lambda\theta_2})]}, \quad \theta_1 < x, \; \theta_2 < y. \tag{3.1}$$

The marginal distribution of X is given by

$$h(x) = \frac{\lambda e^{-\lambda(x-\theta_1)}[1+\alpha(e^{-\lambda\theta_2}-1)(2e^{-\lambda x}-1)]}{1+\alpha(e^{-\lambda\theta_1}-1)(e^{-\lambda\theta_2}-1)}, \quad \theta_1 < x.$$

It can be easily verified that when

$$\frac{1+\alpha}{\alpha}=e^{-\lambda\theta_2}, \tag{3.2}$$

we get

$$h(x)=2\lambda e^{-2\lambda(x-\theta_1)}, \quad \theta_1<x.$$

Also, the joint probability density function of (X,Y) can be written as

$$f_2(x,y)=\lambda^2 e^{-\lambda(x+y-2\theta_1-\theta_2)}[1+\alpha(2e^{-\lambda x}-1)(2e^{-\lambda y}-1)], \quad \theta_1<x, \ \theta_2<y.$$

In this case, the marginal distribution of Y is given by

$$g(y)=\frac{\lambda\alpha}{1+\alpha}e^{-\lambda(y-\theta_1)}[1+\alpha(e^{-\lambda\theta_1}-1)(2e^{-\lambda y}-1)], \quad \theta_2<y.$$

Let (x_i,y_i), $i=1,2,\ldots,n$ be a random sample from the probability density function f_2. Also, let $x_{1:n}<x_{2:n}<\ldots<x_{n:n}$ and $y_{1:n}<y_{2:n}<\ldots<y_{n:n}$ be the corresponding order statistics of x_i's and y_i's respectively.

Using Guenther (1978), UMVU estimators of θ_1 and λ are given by

$\hat{\theta}_1=\frac{nX_{1:n}-\bar{X}}{n-1}$ and $\hat{\lambda}=\frac{n-2}{2n(\bar{X}-X_{1:n})}$, respectively. Hence, with $x_{1:n}=3.30$ and $y_{1:n}=3.27$ (see Table I), we have $\hat{\theta}_1=2.36$ and $\hat{\lambda}=0.0113$.

We now find the moment type estimator of θ_2 using the marginal distribution of Y.

We can easily show that

$$E(Y)=\frac{1}{2\lambda}[1+2\lambda\theta_2-\alpha+e^{\lambda\theta_1}(1+\alpha)].$$

Therefore,

$$37.8384=\frac{1}{2(0.0113)}[1+2(0.0113)\theta_2-\alpha+e^{(0.0113)(2.36)}(1+\alpha)].$$

Using (3.2) and after simplification, we get

$$50.75+\theta_2-\theta_2e^{(0.0113)\theta_2}-49.56e^{(0.0113)\theta_2}=0.$$

Now using Newton's method with 3 iteration, we have $\hat{\theta}_2=1.9895$ and hence $\alpha=-44.98$.

Estimation of the probability that Y is less than has been considered in the literature for many years. This problem arises in the context of reliability studies. A number of papers deal with the estimation when X and Y are independent exponential variates. But, in this study, we derive the estimator of Pr[Y<X] when X and Y are dependent variates.

We can easily show that

$$P[Y<X]=\frac{e^{\lambda(\theta_1+\theta_2)}}{[1+\alpha(1-e^{-\lambda\theta_1})(1-e^{-\lambda\theta_2})]}\Big[e^{-\lambda(\theta_1+\theta_2)}-\frac{1}{2}e^{-2\lambda\theta_1}-\alpha e^{-2\lambda\theta_1}\{-e^{-2\lambda\theta_1}+\frac{1}{2}e^{-3\lambda\theta_1}-e^{-\lambda(\theta_1+2\theta_2)}+e^{\lambda(\theta_1+\theta_2)}+\frac{1}{2}e^{-\lambda\theta_1}+e^{-2\lambda\theta_2}-e^{-\lambda\theta_2}\}\Big].$$

Hence, the estimator of Pr[Y<X] is 0.5045.

Since the measurements are taken few days apart, we now accommodate the decay rate and fit the model using the new data set. Using the decay rate 0.01186, we note that $x_{1:n}$ = 3.152 and $y_{1:n}$ = 3.27. In this case, $\hat{\theta}_1$ =2.25, $\hat{\theta}_2$ =1.93, α =-44.18 and Pr[Y<X]=0.5041.

Next , we fit the model for the data set for which the dioxin concentration is greater than 10 ppt. Note that from table II, $x_{1:n}$ = 12.915 and $y_{1:n}$ = 11.26. In this case, $\hat{\theta}_1$ =11.50 and $\hat{\theta}_2$ =9.95 and Pr[Y<X]=0.5172.

4. CONCLUSION

In this study, we consider two left truncated bivariate exponential parameter density functions. The first model is mainly applicable in system reliability studies. The second model is a modified truncated version of Marshall and Olkin distribution. Under certain condition, it is shown that this distribution has exponential marginal. Lehmann-Scheffè theorem is used to estimate the parameters for the first model. For the second model, we use moment type estimators and UMVU estimators. We also find the probability that Y is less than X. An Air Force Health Study reliability data set is applied to our

results. The graphs show that the second model fit very well for the data set. In the future, a new reliability study could be carried out for the dioxin data using this model.

ACKNOWLEDGEMENTS

I would like to thank the Armstrong laboratory and the Air Force office of Scientific Research for sponsoring this research. This work was done under the guidance of Dr. Joel E. Michalek. The author thanks Dr. Michalek for his assistant in preparing this manuscript.

REFERENCES

Arnold, B. C. and Strauss, D. (1988). Bivariate Distributions with Exponential Conditionals. Amer. Statist. Assoc. J., 83, 522-527.

Beg, M. A. (1980). On the Estimation of Pr[Y>X] for the Two-Parameter Exponential Distribution. Metrika, 27, 29-34.

Guenther, W. C. (1978). Some Easily Found Minimum Variance Unbiased Estimators, Amer. Statist. , 33, 29-34.

Harris, R. (1967). Reliability Applications of a Bivariate Exponential Distributions. J. Am. Stat. Assoc., 18-27.

Johnson, N. L. (1975). Letter to the editor. Technometrics, 17, 393.

Kelly, G. D. and Kelly, J. A. and Schucany, W. R. (1976). Efficient estimation of Pr(Y<X) in the exponential case. Technometrics, 18, 359-360.

Marshall, A. W. and Olkin, I. (1967). A multivariate Exponential Distributions. J. Am. Stat. Assoc., 62, 30-44.

Selvavel, K. (1992). Uniform Minimum Variance Unbiased Estimators for Bivariate Families. Metrika, 39, 131-138.

Tong, H. (1974). A note on the estimation of P(Y < X) in the exponential case. Technometirics, 16, 625.

Table I

Physical Examination Date, the number of Days Between the Dioxin Blood Draws,

Dioxin Levels and Percent Body Fat

in the 46 Ranch Hand Veterans with Paired Dioxin Results

		Dioxin Results (ppt)		Percent Body Fat	
PE Date[1]	Days[2]	First[3]	Second[4]	Tour[5]	1987[6]
26 Oct 87	199	13.50	15.29	15.7	21.5
26 Oct 87	199	16.40	12.69	14.0	16.5
11 Nov 87	215	3.60	4.63	.	16.1
12 Oct 87	185	8.90	10.65	.	17.2
19 Oct 87	192	23.10	12.43	12.0	21.0
04 Nov 87	208	35.50	30.67	18.6	24.4
20 Jan 88	285	10.00	9.70	.	13.8
05 Oct 87	178	24.90	15.82	20.0	26.1
14 Oct 87	187	29.20	23.29	21.4	27.1
05 Oct 87	178	24.50	15.51	15.8	25.0
05 Dec 87	239	42.00	33.21	27.0	31.0
19 Oct 87	192	157.00	82.24	21.5	28.2
26 Oct 87	199	86.70	74.44	20.1	17.7
28 Sep 87	171	79.40	66.37	15.7	21.4
21 Sep 87	164	210.50	144.39	20.7	36.8
16 Nov 87	220	11.30	8.50	15.7	16.3
20 Jan 88	285	3.50	3.80	.	12.1
12 Oct 87	185	65.00	55.05	17.7	19.4
16 Sep 87	159	32.70	21.89	14.4	18.0
04 Nov 87	208	3.30	3.27	16.6	22.6
14 Oct 87	187	48.20	53.46	20.5	28.0
21 Sep 87	164	8.70	7.03	.	26.1
21 Oct 87	194	5.30	6.59	.	15.9
21 Oct 87	194	40.70	58.05	18.7	33.0
19 Oct 87	192	58.30	52.80	19.5	20.3
08 Feb 88	304	9.00	7.90	.	29.9
25 Jan 88	290	131.70	99.59	18.7	20.2
22 Feb 88	318	64.90	39.20	17.5	17.2
16 Nov 87	220	56.10	66.96	12.8	18.8
04 Nov 87	208	9.50	12.71	.	22.0
10 Feb 88	306	23.70	25.72	19.9	24.8
28 Sep 87	171	167.60	141.56	12.9	21.2
14 Oct 87	187	10.80	7.99	15.2	20.6

Table 1 (Continued)

PE Date[1]	Days[2]	Dioxin Results (ppt) First[3]	Second[4]	Percent Body Fat Tour[5]	1987[6]
23 Sep 87	166	24.00	42.53	16.1	21.4
15 Feb 88	311	3.80	4.19	.	14.9
07 Dec 87	241	4.30	3.93	.	17.4
29 Feb 88	325	14.40	11.26	19.3	18.6
16 Sep 87	159	36.10	25.74	11.7	19.3
02 Nov 87	206	44.50	41.21	13.9	16.4
29 Feb 88	325	26.40	22.12	17.5	23.7
16 Nov 87	220	132.90	166.64	16.1	19.8
24 Feb 88	320	26.70	12.97	12.0	13.2
12 Oct 87	185	90.80	57.97	21.4	25.4
06 Jan 88	271	72.40	48.14	14.4	23.0
05 Feb 88	301	78.20	56.22	17.6	29.3
23 Sep 87	166	26.50	24.02	16.7	20.8

1. Date of physical examination (PE). The pilot study date was 10 April 1987 for all subjects.
2. Days between the pilot (first) and PE (second) blood draws.
3. Dioxin result from blood drawn during the pilot study in parts per trillion (ppt).
4. Dioxin result from blood drawn during the 1987 physical examination in ppt.
5. Percent body fat during the subject's tour of duty in Vietnam.
6. Percent body fat during the 1987 physical examination.

Table II

OBS	FIRST	DAYS	SECOND	CFIRST
1	13.5	199	15.29	12.915
2	16.4	199	12.69	15.690
3	23.1	192	12.43	22.134
4	35.5	208	30.67	33.895
5	24.9	178	15.82	23.933
6	29.2	187	23.29	28.010
7	24.5	178	15.51	23.549
8	42.0	239	33.21	39.825
9	157.0	192	82.24	150.436
10	86.7	199	74.44	82.946
11	79.4	171	66.37	76.437
12	210.5	164	144.39	202.959
13	65.0	185	55.05	62.379
14	32.7	159	21.89	31.564
15	48.2	187	53.46	46.236
16	40.7	194	58.05	38.981
17	58.3	192	52.80	55.863
18	131.7	290	99.59	123.473
19	64.9	318	39.20	60.468
20	56.1	220	66.96	53.421
21	23.7	306	25.72	22.141
22	167.6	171	141.56	161.345
23	24.0	166	42.53	23.130
24	14.4	325	11.26	13.396
25	36.1	159	25.74	34.846
26	44.5	206	41.21	42.507
27	26.4	325	22.12	24.559
28	132.9	220	166.64	126.553
29	26.7	320	12.97	24.866
30	90.8	185	57.97	87.139
31	72.4	271	48.14	68.165
32	78.2	301	56.22	73.136
33	26.5	166	24.02	25.539

Figure I

Marginal Distribution of X Using Air Force Reliability Data

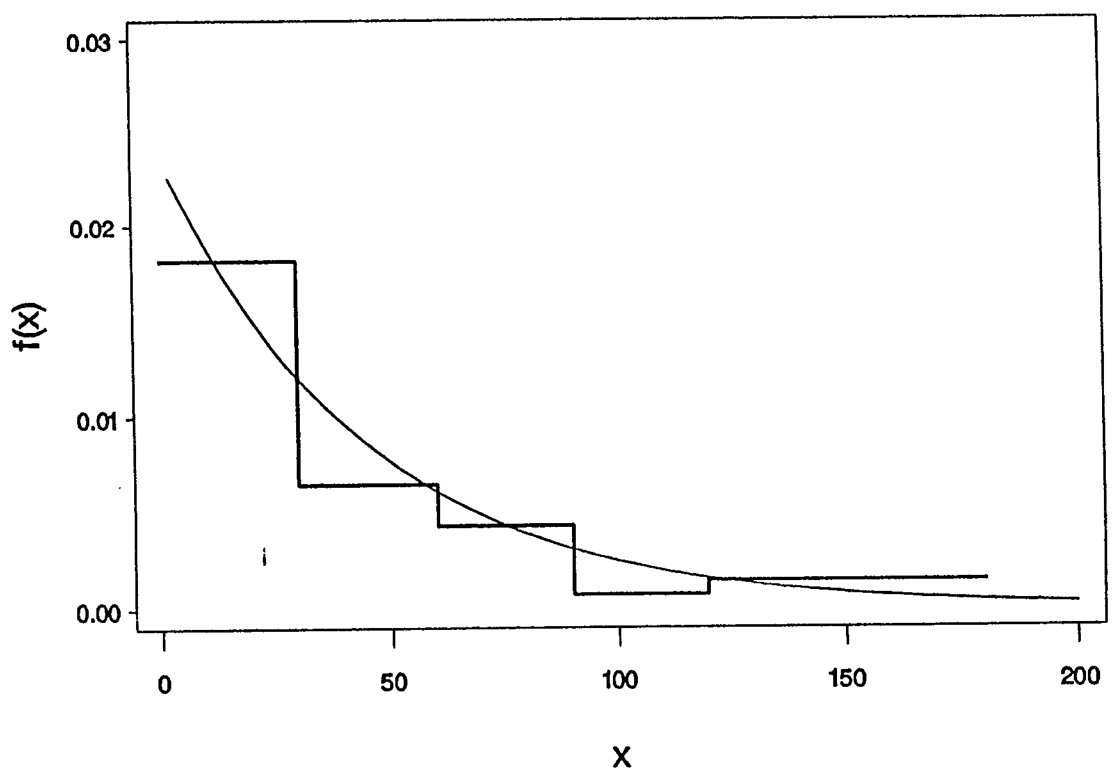


Figure II

Marginal Distribution of Y Using Air Force Reliability Data

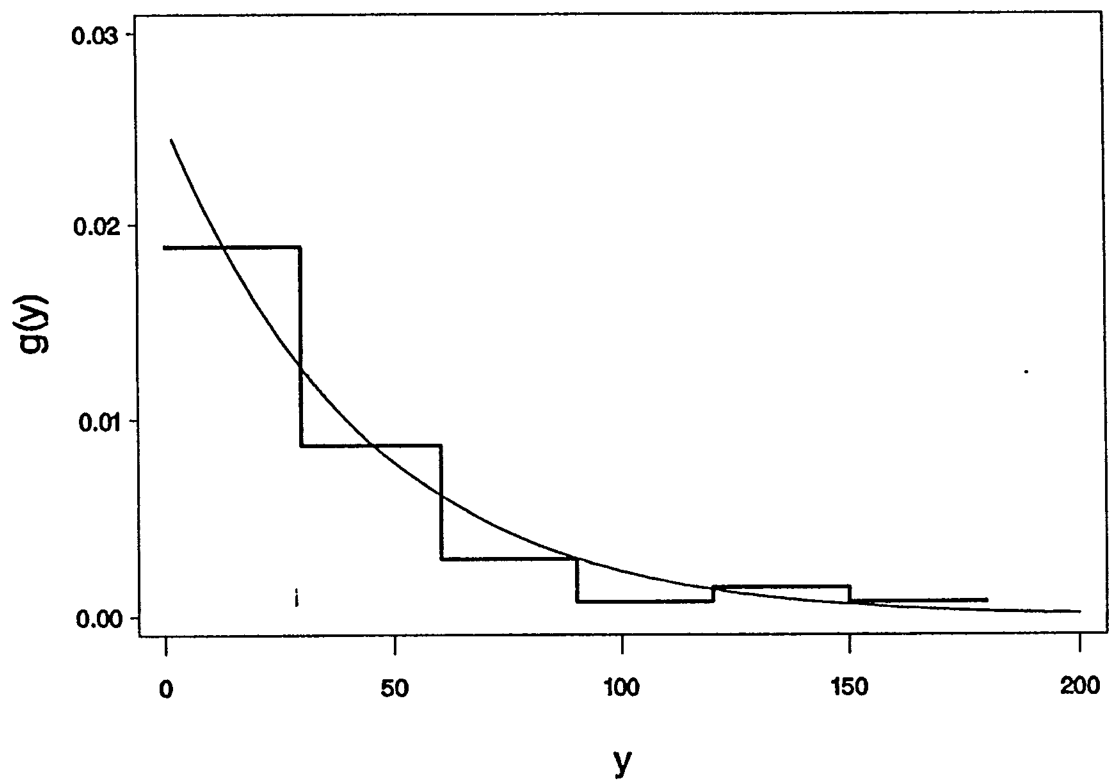


Figure III

Marginal Distribution of X Using Air Force Reliability Data Using Dioxin > 10 and No Decay Rate Model

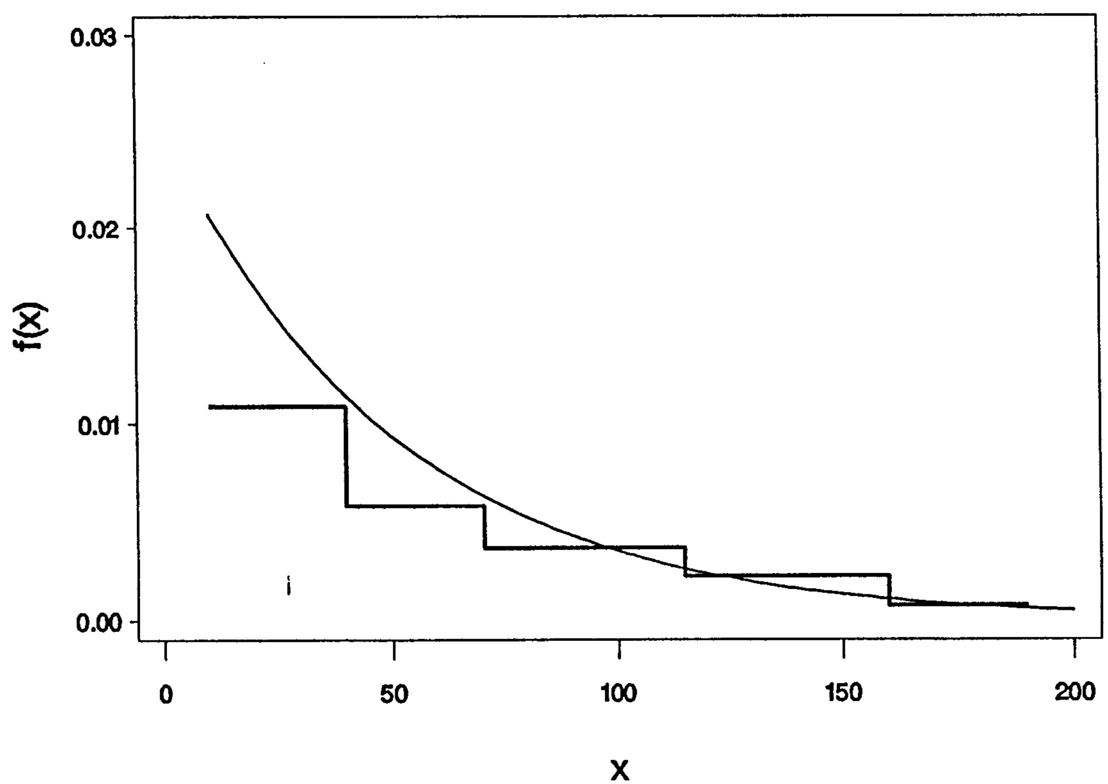


Figure IV

Marginal Distribution of Y Using Air Force Reliability Data Using Dioxin > 10 and No Decay Rate Model

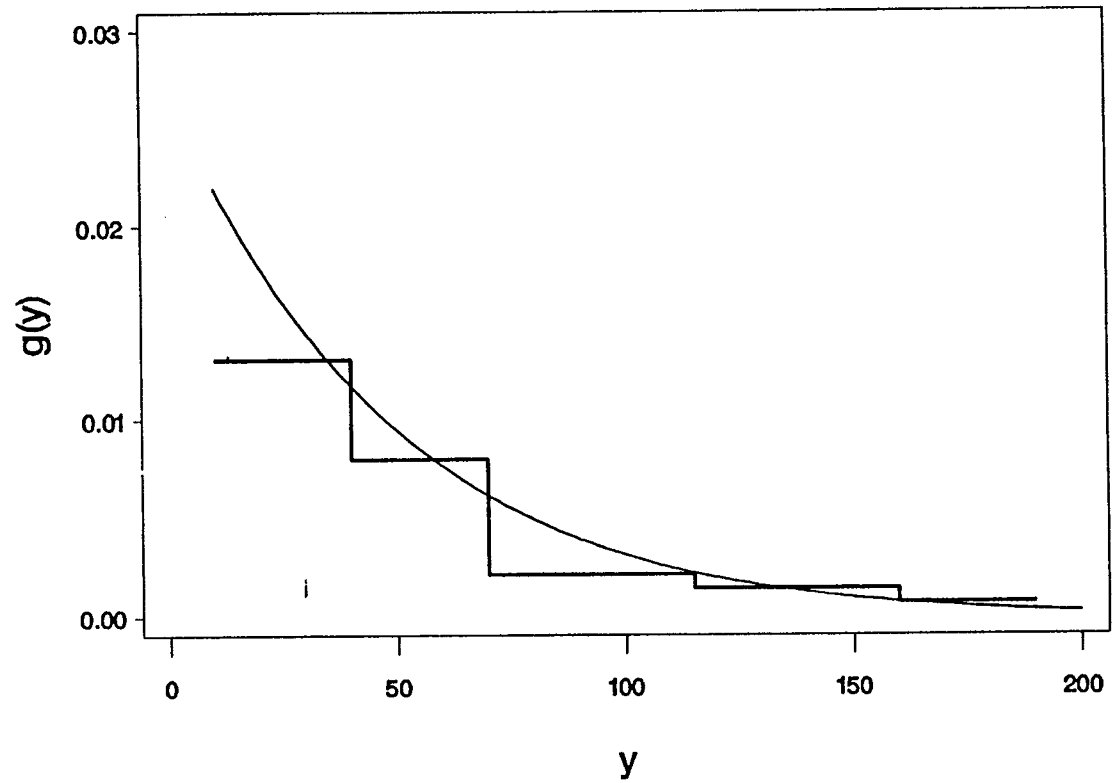
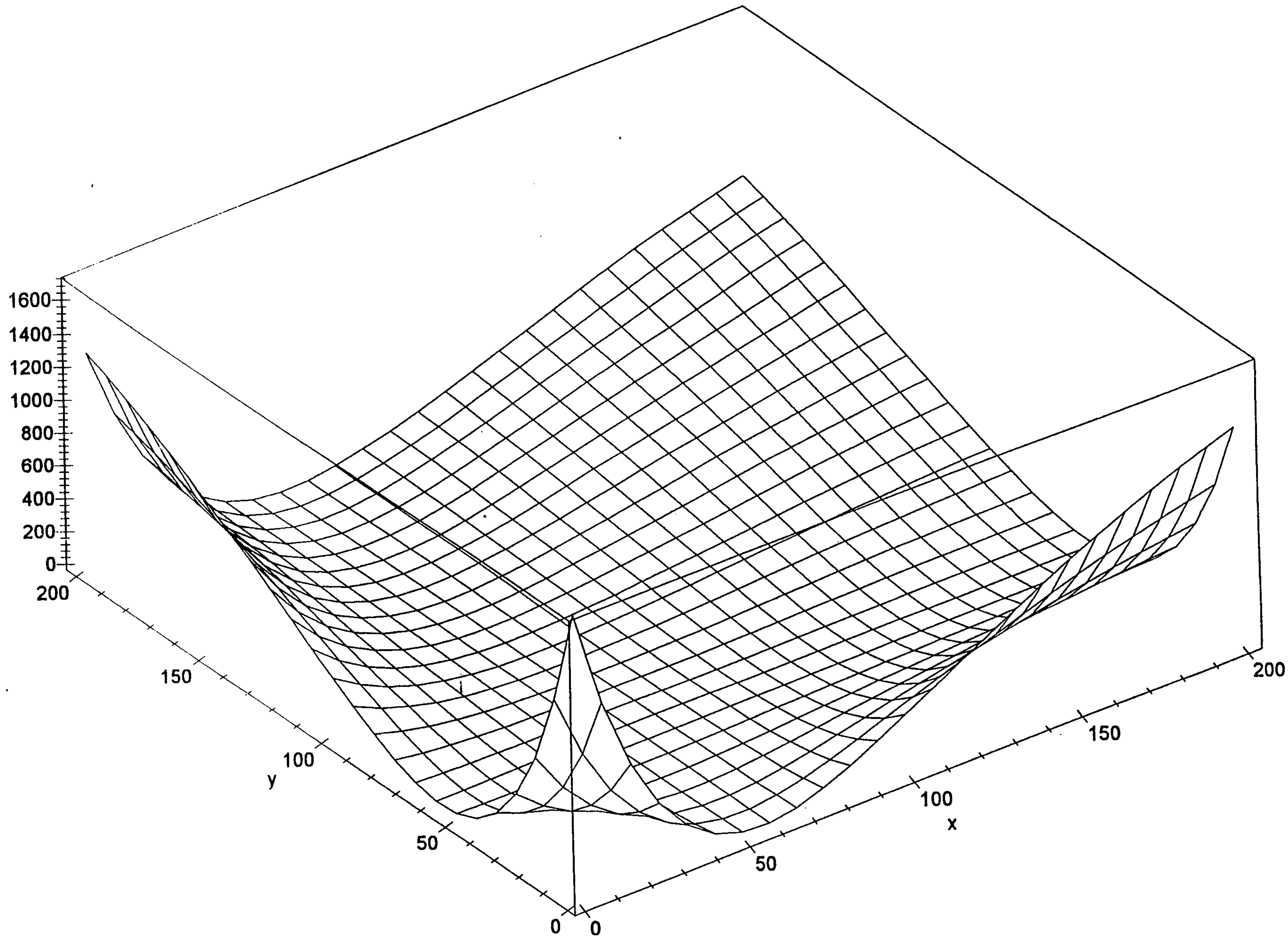


Figure v


1600
1400
1200
1000
800
600
400
200
0
200
150
100
y
50
0
0
50
100
x
150
200

BIODEGRADATION OF 2,4-DNT AND 2,6-DNT IN MIXED CULTURE AEROBIC FLUIDIZED BED REACTOR AND CHEMOSTAT

Barth F. Smets
Assistant Professor
Environmental Engineering Program

University of Connecticut
261 Glenbrook Rd, U-37
Storrs, CT 06268

Final Report for:
Summer Faculty Research Program
Armstrong Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, DC

and

Armstrong Laboratory

September 1996

BIODEGRADATION OF 2,4-DNT AND 2,6-DNT IN MIXED CULTURE AEROBIC FLUIDIZED BED REACTOR AND CHEMOSTAT

Barth F. Smets
Assistant Professor
Environmental Engineering Program
University of Connecticut

Abstract

This study examined the aerobic biodegradation of mixtures of 2,4-dinitrotoluene (2,4-DNT) and 2,6 dinitrotoluene (2,6-DNT). A 1.5 L fluidized bed bioreactor with sand (d: 0.425-0.595 mm) as a carrier material was fed a tap water laden with 2,4-DNT and 2,6-DNT at nominal concentrations of 40 and 10 mg/L, respectively. The loading rates to the reactor were gradually increased by increasing the feed flow rate from 0.12 to 1.0 L/hr. Removal efficiencies higher than 99% for 2,4-DNT and 95% for 2,6-DNT were obtained for applied surface loadings of 240 mg/m^2 d 2,4-DNT and 60 mg/m^2 2,6-DNT. Nitro-N was almost stoichiometrically released and mainly measured as nitrate-N in the reactor effluent. Biofilm concentrations in the reactor at the highest loading rate were 2.22 mg COD/g sand (SD 0.11) and 0.65 mg protein/g sand (SD 0.08). Biofilm thickness was estimated at 45.8 μm (SD 1.9 μm). Air scouring during resuspension of settled bed resulted in significant biofilm washout. Respirometric experiments were successfully applied to measure biotransformation of 2,4-DNT and 4-methyl-5-nitrocatechol (4M5NC) by chemostat cultures. A stoichiometry of 3.59 (SD 0.39) and 2.39 (SD 0.22) moles of oxygen / mole of substrate was measured for 2,4-DNT and 4M5NC, respectively. 2,4-DNT removal followed typical Monod kinetics, while 4M5NC removal exhibited strong substrate inhibition kinetics.

BIODEGRADATION OF 2,4-DNT AND 2,6-DNT IN MIXED CULTURE AEROBIC FLUIDIZED BED REACTOR AND CHEMOSTAT

Barth F. Smets

Materials and Methods

Fluidized Bed Reactor Vessel A 1.5 L volume water jacketed fluidized bed reactor with an inner diameter of 5.2 cm (Bioengineering, Wald ZH, Switzerland) was filled with 0.74 kg of acid washed Ottawa sand (0.425-0.595 mm diameter). The conical part at the bottom of the reactor was filled with 3 mm stainless steel bearing balls to facilitate flow distribution. Temperature in the reactor was measured and maintained at 20 °C by recirculation water through the reactor jacket using a recirculation cooling bath. pH was controlled between 7 +/- 0.1 by automatic addition of NaOH/KOH (1 M each) and phosphoric acid (10% v/v) to the recirculation flow. Aeration was provided using a peristaltic pump delivering lab air to the recirculation line (7.02 l/h for HRT of 3, 1.5, and 0.75 h). Dissolved oxygen concentration was monitored in the top of the reactor bed and maintained higher than 40% of air saturation using a Ingold DO electrode interfaced with a calibrated DO meter (Ingold, Wilmington, MA). Recirculation through the bed was maintained using a centrifugal pump (March Pumpen GmbH, Germany) controlled at approximately 1.5-1.6 L/min. as measured using an in-line flow meter (Gilmont Instruments, Barrington, IL) resulting in approximately 40% bed expansion. The fluidized bed reactor was operated at hydraulic retention time (HRT) of 12, 6, 3, 1.5 h in this sequence. For each HRT, steady-state was assumed to be achieved when the concentrations of 2,4- and 2,6-DNT were varied less than 20% over a period of 3 days.

FBR feed was prepared in 150 L batches in stainless steel barrels and consisted of 40 mg/L of 2,4 DNT, 10 mg/L 2,6 DNT, and 70 mg/l H_3PO_4 in tap water. Feed was delivered using peristaltic pumps using Pharmed® pump tubing while feed lines consisted of stainless steel and glass. An in-line placed graduated 10 or 20 ml pipette allowed for daily measurements of flow rates. An in line sample vessel allowed to collect influent samples just prior to entry in the reactor. Feed was delivered to the FBR via the recirculation line. To minimize growth in feed lines, they were rinsed with EtOH and deionized H_2O before a new feed batch was used.

FBR Daily measurements consisted of DO concentration, pH, temperature, bed-height, feed flow rate, recirculation flow rate. Samples were removed to measure concentrations of DNTs, nitrite, and nitrate.

DNT analysis 700 µl of sample was withdrawn from the culture and mixed with 300 µl of methanol in a microcentrifuge vial. Solids were removed by centrifugation at 14,000 rpm for 2 min. and the supernatant was transferred into 2 ml borosilicate glass vials for HPLC analysis. Time between sample withdrawal and full speed centrifugation was less than 20 sec The concentrations of nitrotoluenes was measured on a Hewlett-Packard 1050 HPLC system equipped with an quaternary pump and a variable wavelength-detector. (Hewlett-Packard, Wilmington, DE). Separation of 2,4-DNT, 2,6-DNT, 2,4,6-TNT and 4-methyl-5-nitrocatechol was achieved on a Spherisorb Hexyl (C6) column (Alltech, Deerfield, Il) using a guard column of the same material. 70% H2O, 30% Methanol, acidified with trifluoroacetic acid (0.1%) was used as eluent at a flow rate of 1 ml/min. Compounds were detected at 254 nm.

Nitrite and nitrate analysis 1 ml of culture liquid was centrifuged at 14,000 rpm for 2 min. and the supernatant was transferred into 0.5 ml Dionex Poly-vials (Dionex, Sunnyvale, Calf.). Ions were measured with a Dionex DX-300 Series Chromatography System equipped with a CDM-2 conductivity detector. Separation was achieved on a Dionex AS11 Column. A 19 mM NaOH solution at a flow rate of 0.65 ml/min. served as eluent.

Biofilm COD measurements. With a 0.5 ml sample thimble aliquots of the bed were removed at different depths (top 5 cm, and at half depth). The sample was transferred to one or several COD vial (0-150 or 0-1500 PPM range, HACH). Sample thimbles were rinse with di H_2O. Additional di H_2O was added to the COD vials to a final volume of 2 ml and vials were digested @ 150°C for 2 hours. The COD concentration was read using a HACH spectrophotometer and appropriate potassium phthalate standards. Controls COD vials were included that contained clean acid washed sand. Vials were opened, the supernatant discarded, and the sand was rinsed with di H_2O to remove all reagent. Vials were transferred to 102 °C oven and dried overnight. Vials were brought to room temperature in a dessicator and weighed on an analytical balance. Then sand was removed by inverting and vials reweighed. The biomass COD concentration was calculated as mass of COD/mass of sand. A minimum of 5 COD and sand DW measurements were made at each steady state.

Protein measurement Protein concentrations in the FBR-effluent and on the sand were measured. 3 x 1 ml aliquots of FBR - liquid were transferred into 5 ml borosilicate glass vials. 0.5 ml of bed material (sand with biofilm) was distributed into a series of preweighed 5 ml vials (4 to 12 depending on the expected protein concentration). To these vials water was added to give 1 ml of liquid together with the sample. Blank controls with fresh uncolonized sand and water were measured in triplicates along with bovine serum albumin (BSA) standards at concentrations between 0 and 100 μg/ml. To all vials 1 ml of BCA protein reagent was added (Pierce, Rockford, Ill.). In order to remove the biofilm from the sand vials were vortexed for 20s and subsequently sonicated for 2 min. This procedure was carried out twice initially, and once after 15 min. of incubation at 30 °C. After 30 min. incubation all vials were vortexed and the supernatant of the vials containing sand was transferred into 1.5 ml microcentrifuge tubes and centrifuged at 14,000 rpm for 2 min. in order to remove suspended solids. Absorbance was measured at 562 nm and protein concentrations were calculated with the BSA standard curve. The amount of sand was weighed after the vials were carefully washed 5 times (without loosing sand) with deionized water and dried overnight at 104 C. Protein concentration in the liquid are given in mg/l and in the biofilm attached to the sand in mg/g of sand.

Biofilm density measurements With a 0.5 ml sampling thimble a first aliquot from bed 5 cm from the top of bed and a second aliquot at approx. half bed depth were removed. Free water floating on top of the thimble was removed using a KimWipe®. The contents of one thimble was transferred to four different aluminum weighing pans using a small spatula and subsequently weighed on an electronic balance. Pans were transferred to 102 °C oven and dried overnight. Pans were brought to room temperature in a dessicator and weighed on an analytical balance. Approximately 200 μl of diH_2O was added to each pan and the sand transferred to one or several COD vials (0-150 or 0-1500 PPM). Pans were rinsed with H_2O to redissolve any non-attached biofilm that stuck to the aluminum pan. Additional di H_2O was added to the COD vials to 2 ml total. COD and sand dry mass concentrations were determined as described above. The average weight of evaporated water (W), the total mass of dry sand (M), and the mg XCOD/g sand were calculated corresponding with the 0.5 ml sample removed from the bed. The biofilm thickness was calculated using the following equation derived from Rittmann *et al.* (1986), where a/m is the surface area/mass of sand, which was calculated as 5.35 m^2/kg, and ρ = density of water. This equation assumes that the native biofilm consist for 0.99% out of H_2O:

$$L_f = \frac{W}{M*(a/m)*\rho*0.99}$$

FBR Shift-Load experiments The FBR was subjected to short-term (4 hour) shifts in the applied surface loadings by varying the feed flow rates. When the FBR attained steady-state, the flow rate through the reactor was instantaneously increased or decreased (from 0.25 to 2.0 times the steady-state flow rate). Flow rates were monitored during the transient load experiments, and samples for HPLC analysis were periodically removed from the top of the reactor bed. After 4 hours the reactor was returned to steady-state flow rate for at least 8 hours prior to subjecting the FBR to another transient load experiment. Stop-flow experiment were also performed whereby the feed to the reactor was instantaneously stopped and the DNT concentration in the reactor monitored. All transient experiments were performed typically within a 3-day period at a given steady state.

Physical Calculations on FBR. Physical calculations on the FBR, specifically aimed at estimating the mass transfer layer for 2,4 and 2,6-DNT external to the biofilm are in Appendix 1. The calculations reveal that, under with the given operating parameters, the average thickness of the external mass transfer layer is approximately 20 μm.

Chemostat Operation. A 2-liter chemostat reactor (Virtis, Omni-Culture, Virtis, Gardiner, NY) was operated with a working volume of 1.6 L, 0.2 μm-filtered (Gelman Sciences, Ann Arbor, MI) lab air was provided at a flow rate of 0.2 L/min. which resulted in a dissolved oxygen concentration of at least 7.5 mg/L as monitored by a Ingold DO probe, reactor contents were stirred at 400 rpm, and temperature was controlled at 30 ± 1 °C using a heater element and recirculating coolant element.

Chemostat Feed The reactor was fed with a medium containing 2,4-DNT and 2,6-DNT as sole carbon sources at a target total DNT concentration of 100 mg/L. The first steady state was established at 80 and 20 mg/L and the second steady state at 50 and 50 mg/L of 2,4-DNT and 2,6-DNT, respectively. The mineral salts medium (MSB)

was derived from the Stanier's mineral salts medium (Stanier *et al.*, 1966) with the phosphate buffer added at 50% and all other compounds at 10% of the full medium strength.

Chemostat Daily Measurements were performed as was done for FBR. In addition, biomass concentrations in the chemostat were determined by difference from COD analysis on total reactor mixed liquor and mixed liquor supernatant after a 5 minute centrifugation at 15,000 rpm in a Microcentrifuge (Brand).

Extant Respirometry Respirometric experiments were performed essentially as described by Ellis *et al.* (1996). To concentrate biomass and to avoid the impact of metabolic products present in the culture liquid, 50 ml of the culture suspension was harvested, centrifuged at 15,000 g at room temperature, decanted and resuspended in 5 ml of prewarmed MSB. Cell suspensions were gently aerated using Pasteur pipettes until ready for use. The biomass concentration was determined on an aliquot of the suspension using 0-150 PPM HACH COD vials. The workstation consisted of two water-jacketed 2 ml oxygraph units maintained at 30 ^{0}C using a recirculating water bath. The DO probes were YSI 5331 probes (YSI Yellow Springs, OH), equipped with YSI 5776 high-sensitivity membranes that were replaced for each experiment, connected to a YSI 5300 Biological Oxygen Monitor connected via an interface block to a A/D data-acquisition board (PCL 711, Advantech) which was operated using the Labtech® Notebook® software on a Pentium PC. DO probes were equilibrated and calibrated at 30 ^{0}C in deionized water (DO_{sat} = 7.559 mg/L). Data acquisition was performed at 10 Hz, and data were averaged over a 2 or 4 second interval. Subsequent data analysis was performed on data sets that retained one data point every 2 or 4 seconds. Approximately 2 ml of cell suspension was transferred to each oxygraph chamber, which was subsequently fit with a hollow-core glass stopper, and the absence of bubbles in the chamber was ensured. The suspension was mixed vigorously using a micro stir bar and a stir plate. At least 3 minutes of background oxygen uptake data were collected before substrate injections were made, and cell suspensions were reaerated whenever needed. Microvolume injections were made through the center of the glass stopper from stock solutions of 2,4-DNT (100 PPM) and 4-methyl-5-nitrocatechol (4M5NC) (200 PPM). Initial volumes in oxygraph chamber and volumes added or removed were recorded to calculate effective biomass concentrations and initial substrate concentrations during the individual experiments. Initial substrate concentrations were then expressed in COD units using the conversion factors 1.41 mg COD /mg 2,4-DNT and 1.42 mg COD /mg 4M5NC.

Results

Figures 1, 2, and 3 show the FBR performance in terms of effluent concentrations of 2,4 DNT, 2,6 DNT, and TNT. The reactor was operated at a nominal feed flow rate of 240 ml/h from day 1 through 14, of 480 ml/hr from day 14 through day 35, and of 1050 ml/h from day 68. During the period from day 40 through day 57 several upsets occurred which caused unstable reactor performance; twice significant contamination of the feed bottle resulted in very low influent concentrations to the reactor; malfunctions in the pH controlling resulted in occasional pH upsets ; and unintentional settling of the bed resulted in significant biofilm washout during resuspension on day 53. Hollow symbols in Figures 1 and 2, indicate that during the majority of the operation, fairly steady feed barrel concentrations of 35 mg 2,4-DNT and 10 mg/L 2,6 DNT were maintained. Throughout the study removals exceeded 99% and 95% for 2,4-DNT and 2,6-DNT, respectively, during steady-states. Trace quantities of TNT were present in the feed because of impurities in the 2,4-DNT chemical stock. Significant removal of TNT (> 70%) was achieved throughout the study as well, as indicated in Figure 3.

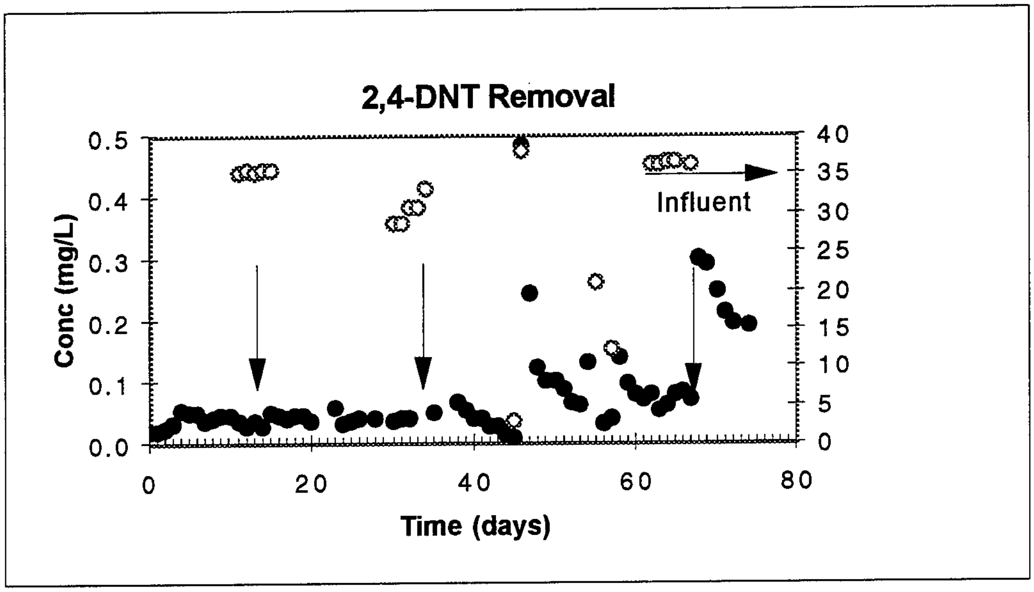


Figure 1. 2,4-DNT influent and effluent concentrations during FBR operation
Vertical arrows indicate the times when the applied loading rates were changed.

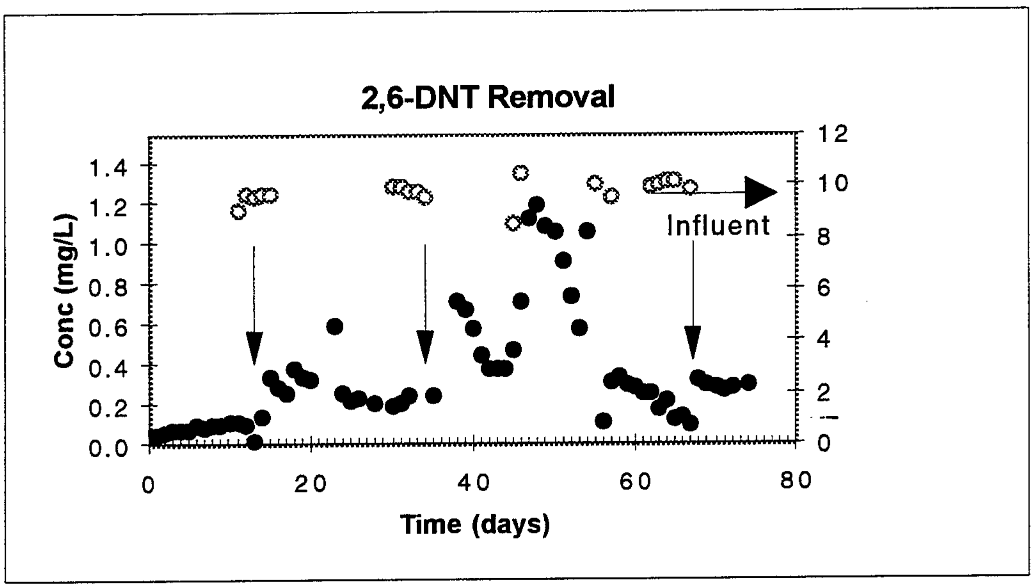


Figure 2. 2,6-DNT influent and effluent concentrations during FBR operation
Vertical arrows indicate the times when the applied loading rates were changed.

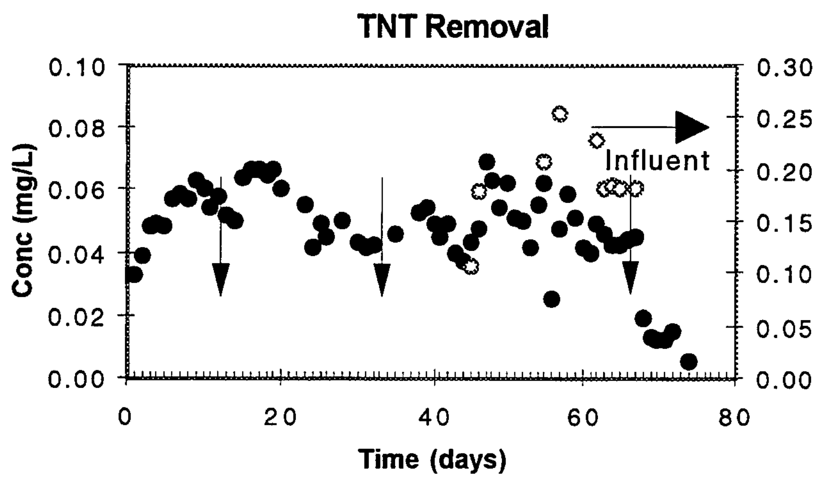


Figure 3. TNT influent and effluent concentrations during FBR operation

Performance in terms of 2,4-DNT and 2,6-DNT are presented in terms of the applied surface loading rates (mass of DNT applied/area of biofilm surface/unit time) in Figure 4. The applied surface loading was calculated as

$$J = \frac{FS_0}{M*(a/m)}$$

with F = feed flow rate (L/day), S_o = influent concentration (mg/L), and other symbols are as defined before

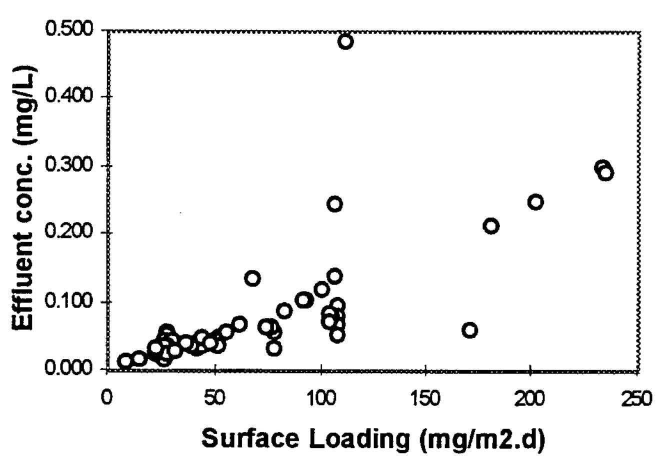


Figure 4. 2,4-DNT effluent concentrations as a function of the applied 2,4-DNT surface loading

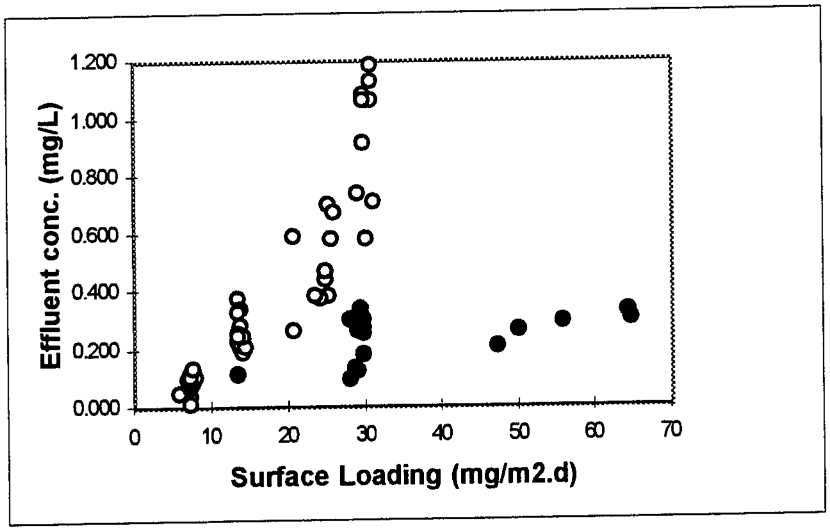


Figure 5. 2,6-DNT effluent concentrations as a function of the applied 2,6-DNT surface loading Hollow symbols refer to data prior to biofilm washout; filled symbols refer to data after biofilm washout.

For 2,4-DNT effluent concentrations increase monotonically with increases in the applied surface loading rates. The data suggest that the FBR was never operated in a true low load region, where increases in the load do not impact the effluent concentration. Loadings as high as 100 mg/m^{2}d resulted in effluent concentrations in compliance with the MCL of 200 μg/L. The results for 2,6-DNT removal can be interpreted when the data prior to and after the biofilm loss incident on day 53 are plotted separately as shown in Figure 5. Prior to the biofilm loss, effluent concentrations increase with increases in surface loading, as expected. A significant change in the physiology of the retained dominant 2,6-DNT degrading population occurred following the biofilm loss incident. At a loading of 30 mg/m^{2}d, effluent concentrations dropped from 0.6 to 1.2 mg/L to the 0.1 to 0.4 mg/L range. After the biofilm loss, removal of 2,6-DNT was clearly superior.

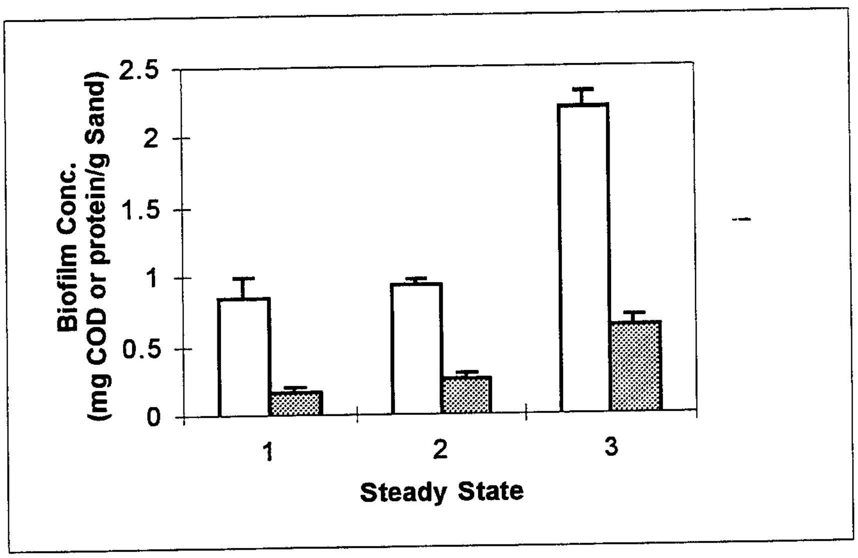


Figure 6. Steady-state biofilm concentrations at the three different steady-states in mg/g sand. Open symbols represent COD units, shaded bars represent protein concentrations.

Figure 6 lists the biofilm concentrations (per mass of sand) that were measured at different steady state during the study. Although the biofilm concentration increase from the first to the second steady state is small, it is very

apparent with the second 2-fold increase in the applied sludge loading. Biofilm density and biofilm thickness were determined during the last steady state: $L_f = 45.8\ \mu m$ (stdev 1.9 µm), $\rho_X = 8210\ g/m^3$ (stdev: 908 g/m^3).

The effluent concentrations of the examined N-species are illustrated in Figure 7. Two important observations can be drawn from the data. First, the majority of the effluent nitro-N is in the form of NO_3^--N, although nitro group cleavage from DNT is expected to yield NO_2^--N . This suggest that an active population of NO_2^--N oxidizing nitrifiers were present in the bioreactor. Second, the measured effluent nitro-N concentrations are not significantly lower than the calculated effluent nitro-N concentrations. This suggest that other biochemical reactions which remove NO_2^--N and NO_3^--N are not very important. It should be noted that the calculated effluent nitro-N concentration is a very conservative estimate. Indeed, some of the released nitro-N must be used for cell synthesis during DNT biodegradation and during nitrification. However, because the growth yield of the latter two processes could not be determined in the FBR, the importance of nitro-N removal via those mechanisms could not be estimated. As a result, the theoretical effluent nitro-N concentration should be even less than that depicted in Fig. 7. This analysis does, therefore, confirm that other nitro-N converting routes, such as denitrification, were of limited importance in the FBR.

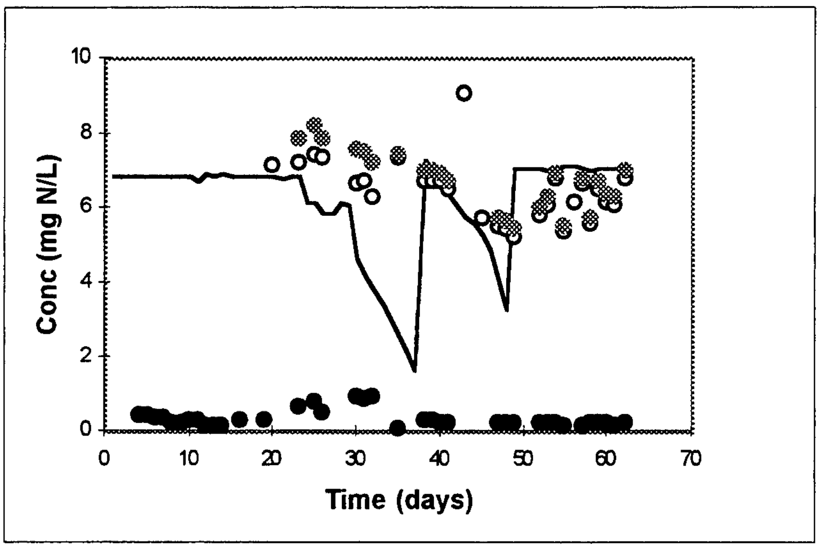


Figure 7. NO_2^--N and NO_3^--N effluent concentrations (mg/L)during the FBR operation. Hollow circles: NO_3^--N, black-filled circles: NO_2^--N; gray-filled circles: NO_2^--N + NO_3^--N. The continuous line is the calculated effluent NO_2^--N + NO_3^--N concentration based on influent and effluent DNT concentrations assuming stoichiometric release of NO_2^- -N from DNT.

Earlier batch experiments on 2,4-DNT mineralization by mixed cultures derived from the chemostat revealed the transient accumulation of 4M5NC in the growth medium. This suggested that 4M5NC transformation may be the rate limiting step in 2,4-DNT mineralization. Furthermore, in studies with purified 4M5NC monooxygenase, Haigler and Spain (1996) demonstrated severe substrate inhibition to the enzyme. Prediction of the maximum 2,4-DNT shock load capacity of a culture will thus depend on the kinetics of the limiting 4M5NC transformation. To that effect, respirometric experiments were performed with both 2,4-DNT and 4M5NC as sole substrates.

The stoichiometry of oxygen consumption was determined by monitoring oxygen consumptions of sequential injections of 2,4-DNT (from 1.45 to 5.5 mg/L as COD) and 4-methyl-5-nitrocatechol (from 1.45 to 6.00 mg/L as COD) in oxygraph chambers containing concentrated chemostat cell suspensions. The data were converted to calculate the molar O_2/substrate ratio as tabulated in Table 1.

Table 1. Oxygen Stoichiometry Exhibited by Whole Cells grown in Chemostat Culture

Substrate	O_2/Substrate Mean (Stdev)	# replicates
2,4-dinitrotoluene	3.95(0.39)	12
4-methyl-5-nitrocatechol	2.39(0.22)	10

Two key points can be derived from Table 1. First, the respirometric assays using whole cell suspensions do not merely measure the activity of the initial oxygenases in the 2,4-DNT biodegradation. If only oxygenase activity were measured, a maximum stoichiometric need of 3 moles of O_2 per mole of 2,4-DNT would be expected (for the activity of the 2,4-DNT dioxygenase, the 4-methyl-5-nitrocatechol monooxygenase, and the 2,4,5-trihydroxytoluene dioxygenase). Thus, the respirometric assays appear suitable to measure the stoichiometry of 2,4-DNT mineralization. Second, measurement of 4-methyl-5-nitro-catechol removal, an intermediate in the aerobic 2,4-DNT mineralization pathway, was also possible. The oxygen requirement for 2,4-DNT to 4M5NC conversion could be calculated by difference as 1.56 (0.32) which is more than the expected value of 1 needed for activity of the 2,4-DNT dioxygenase.

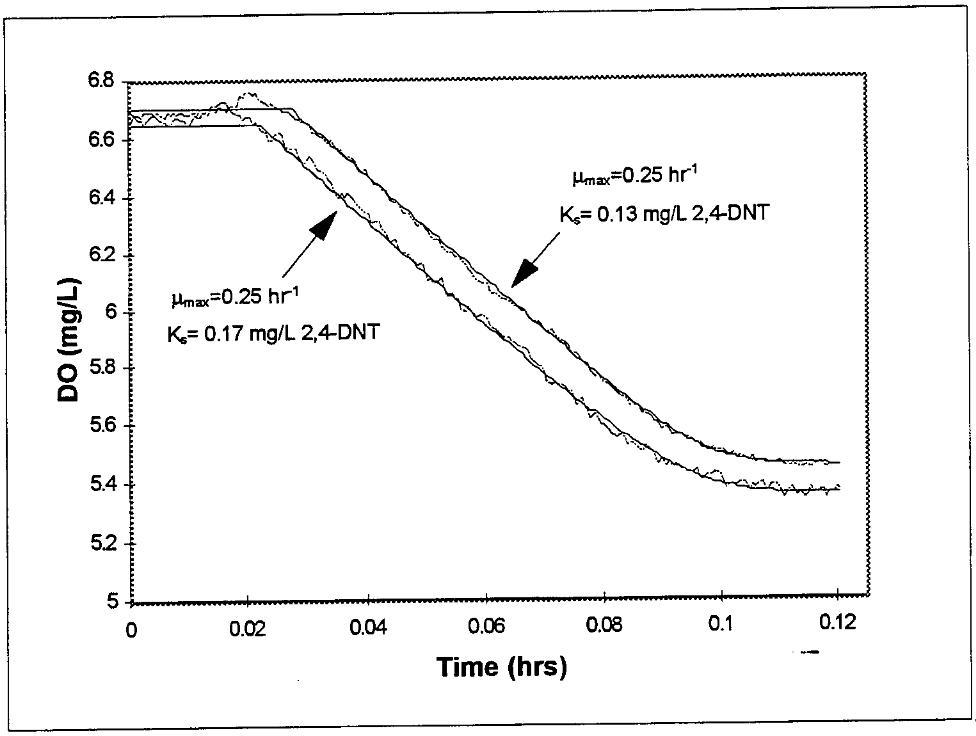


Figure 8. Respirometric response to 2,4-DNT injection of 2.86 mg/L COD in duplicate oxygraph vessels. The biomass concentration in each vessel was 155 mg/L COD. For both cases, the jagged line represents the 2-sec averaged raw data, while the smooth line represents the best fit curve yielding the displayed kinetic parameters.

Respirometric experiments also permitted estimation of the whole cell kinetic parameters for 2,4-DNT and 4M5NC mineralization. Initial 2,4-DNT injection concentrations higher than 2.8 mg/L (4.0 mg/L as COD), yielded biphasic oxygen uptake curves that could not be fit with simple Monod or Andrews kinetic equations. As a result the tabulated kinetic parameters are for 2,4-DNT and 4M5NC initial concentrations of approximately 2 mg/L. Figure 8 illustrates the oxygen uptake profile in duplicate oxygraph vessels containing aliquots of the same chemostat suspension after injection of 2.03 mg/Lof 2,4-DNT (2.90 mg/L as COD). The experimental reproducibility is evident and the data could be fit adequately using a Monod kinetic expression as illustrated by the best-fit curves.

Table 2. Average kinetic parameters for 2,4-DNT and 4M5NC mineralization

Substrate	μ_{max}[&] (hr^{-1})	K_s (mg/L as COD)	K_i (mg/L as COD)	# replicates
2,4-dinitrotoluene	0.25 (0.01)	0.21(0.03)	n/a	4
4-methyl-5-nitrocatechol*	0.99 (0.34)	1.40 (0.45)	1.86 (1.92)	4

[&] Specific growth rates are computed on the basis of the total biomass concentration in the chemostat sample.
* The kinetic parameter estimates for 4M5NC are preliminary. They were obtained using the Solver routine in EXCEL and need to be confirmed using a FORTRAN non-linear parameter estimation routine.

In comparing the kinetics of 2,4-DNT and 4M5NC mineralization two points are evident. First, 2,4-DNT mineralization could be described well using the Monod equation, while 4M5NC mineralization could only be described using a substrate inhibition model such as the Andrews equation (also known as the Haldane equation). Second, self inhibition by 4M5NC is very high. The K_i/K_s ratio for 4M5NC is close to 1 indicating that 4M5NC removal is rapidly inhibited by the 4M5NC concentration and rates close to μ_{max} cannot be achieved.

References

Ellis, T. G., D. S. Barbeau, B. F. Smets and C. P. L. J. Grady. 1996. Respirometric technique for determination of extant kinetic parameters describing biodegradation. Wat. Environ. Res. 68:917-926.

Haigler, B. E. and J. C. Spain. 1996. J. Bacteriol. *In Press*

McCabe, W. L., J. C. Smith and P. Harriot. 1993. Unit Operations of Chemical Engineering. McGraw-Hill, Inc., New York, NY.

Rittmann, B. E., L. Crawford, C. K. Tuck and E. Namkung. 1986. *In situ* determination of kinetic parameters for biofilms: Isolation and characterization of oligotrophic biofilms. Biotechnol. Bioeng. 28:1753-1760.

Stanier, R. Y., N. J. Palleroni and M. Doudoroff. 1966. The aerobic pseudomonads: A taxonomic study. J. Gen. Microbiol. 43:159-171.

Acknowledgement

I wish to express to my gratitude to the the Air Force Office of Scientific Research for providing me with financial assistance in terms of a Summer Faculty Research Award during the period of this work. I also wish to thank colleagues in the Microbiology in the Environics Lab for their input and support and especially Dr. Urs Lendenmann for his collaboration in the reported project.

Appendix A

Hydrodynamic + Physical Calculations on FBR

Bed filling of Bioengineering Fluidized Bed Bioreactor	
Density of Ottawa sand [kg/l]:	1.5
Density of Quarz [kg/m3]:	2200
Diameter of sand fraction [mm]:	0.425-0.595
Dry bed volume [l]:	0.45
Dry bed weight [kg]:	0.74
Sphericity of Sand []:	1
Avarage sand radius [m]:	0.000255
Average grain surface [m2]:	8.17128E-07
Average weight of grain [kg]:	1.52803E-07
Number of grains in bed:	4842837.335
Total surface [m2]:	3.957219251
Total Volume of liquid in reactor (incl. recirc lines) [l]	1.5

Calculation of External Mass Transfer Layer Thickness, L

liquid density (kg/m^3)	ρ	1000			
particle diameter (m)	dp	5.00E-04			
superficial velocity (m/hr)	υ	42.37853514	flow rate (l/min)	F	1.5
abs viscosity (kg/m*sec) or (Pa s)	μ	9.93E-04	reactor ID (m)	din	0.052
porosity	ε	0.61	porosity w/o expansion	εo	0.45
			expansion (%)		140
Diffusion Coeff (m^2/sec)	D	6.35E-10	Estimated from Wilke Chang Equation,		
Re #=	{2ρdpυ/((1-ε)μ)}	30.17585177	(Mc Cabe and Harriot, 1993)		
Schmidt #	μ/ρD	1563.779528			
L(m)=External MT Layer	D*Re^0.75*Sc^0.66/5.7*υ	1.64154E-05			
L(μm)		16.41536076			

A STUDY OF APOPTOSIS DURING LIMB DEVELOPMENT

Mary Alice Smith, PhD
Assistant Professor
Environmental Health Sciences

University of Georgia
206 Dairy Science Building
Athens, GA 30602-2102

Final Report for:
Summer Faculty Research Program
Armstrong Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, DC

and

Armstrong Laboratory

September 1996

A STUDY OF APOPTOSIS DURING LIMB DEVELOPMENT

Mary Alice Smith, PhD
Assistant Professor
Environmental Health Sciences
University of Georgia

Abstract

Apoptosis, or programmed cell death, is believed to be an important component in pattern formation during development. Until the last few years, apoptosis could only be identified by histochemical staining followed by a pathologist's diagnosis. This labor intensive process prevented progress in establishing the location of apoptosis during development and mechanisms responsible for apoptosis. The objectives of this study were to 1) refine an automated procedure for detecting cells undergoing apoptosis, and check the reliability of the procedure by comparing it to the histochemical staining procedure, and 2) to describe the location and amount of apoptosis in the developing mouse limb bud after the dam had been treated with all-trans retinoic acid.

The results of objective 1 ware used to draft a manuscript which is included as a part of this report. The abstract and text of the manuscript follows. Currently the manuscript is in internal review in the Toxicology Division of Armstrong Laboratories, Wright-Patterson AFB after which it will be submitted for publication in the *Journal of Histotechnology.* Data collection for objective 2 is still underway, and a collaboration will be continued to complete the project. An abstract of work completed thus far follows the manuscript.

DETECTION OF APOPTOSIS USING AN AUTOMATED IMMUNOHISTOCHEMICAL PROCEDURE.

MA Smith[1], MC Gothaus[2], A Warren[3], and JR Latendresse[4]. [1]Environmental Health Sciences, University of Georgia, Athens, GA, [2]The Medical College of Ohio, Toledo, OH, [3]Toxicology Division, Armstrong Laboratory, and [4]Mantech Environmental Technology, Inc, Wright-Patterson AFB, OH

Apoptosis, or programmed cell death, occurs during embryonic development, normal tissue homeostasis, oncogenesis and as a result of toxic insult. Understanding the role of apoptosis during these processes is important in learning more about the mechanisms of these normal and abnormal processes. Several assays have been developed to label cells undergoing apoptosis, but the assays are labor intensive and sometimes lack in reproducibility. We describe a 76 step automated procedure which labels apoptotic cells in 4 μm, formalin fixed tissues. The automated procedure is based on the use of capillary action, and uses a commercially available peroxidase-based kit, ApopTag™ by ONCOR (Gathersberg, MD) to do enzymatic end-labeling of DNA fragments and immunohistochemical detection of apoptosis. This procedure allows the processing of 20-40 slides in a single run, reducing interassay variability, saving reagents and reducing technical time. In four different tissue types (liver, mammary gland, uterus and limb bud) apoptotic cells were labeled illustrating the general applicability of the automated procedure. Approximately 85-95% of cells or cellular fragments that were morphologically consistent with apoptotic cells stained golden-brown immunohistochemically with the chromogen, diaminobenzidine. Automation of *in situ* detection of apoptosis is a potentially valuable research tool that can improve the reproducibility of experimental results, conserve the technologist's time and effort, and reduce the quantity of reagents required.

Keywords: apoptosis, immunoperoxidase, immunohistochemistry, capillary gap technology, automation, *in situ* DNA hybridization

Introduction

Apoptosis is a regulated form of physiological cell death that is known to occur during embryonic development (1), immune system maturation (2), normal tissue homeostasis (3,4,5), hormone deprivation of endocrine sensitive cells (6,7),metabolic stress (8), and oncogenesis (9,10). Recently, research has focused extensively on understanding the processes by which apoptosis occurs and its function in the regulation of cell growth. Apoptosis is known to be an active process, probably requiring gene transcription and protein synthesis (8, 11). The series of molecular events triggering and the morphological characteristics of

apoptosis or programmed cell death are thought to be similar across species and cell types (8, 12).

In ours and other laboratories, qualitative and quantitative morphological evaluation of apoptosis during exposure and post-exposure to various chemicals and materials are important for the interpretation of toxicological data, and for understanding mechanisms by which certain xenobiotics affect cell growth. Moreover, the determination of the relationship that exists between cell proliferation and apoptosis using various immunohistochemical techniques can lead to a better understanding of the mechanism of action of various xenobiotics. Studying these relationships requires the analysis of many tissue samples requiring numerous laborious assay steps and providing ample opportunity for experimental error due to sample processing.

Commercially available kits used to detect apoptosis *in situ* are currently available, and nucleotidyl labeling techniques have successfully localized internucleosomal double-stranded DNA breaks that are known to occur in high concentrations during this type of cell death (7,10,12). The *in situ* procedure has the advantages over electrophoretic detection of apoptosis in preparations derived from tissue homogenates by elucidating the precise localization and identification of individual cells undergoing programmed death. The quantitation of apoptotic cells with the assistance of computer-based image analysis requires the optimization of steps that minimize procedural variability. Automated *in situ* detection techniques using microcapillary gap technology allow for the rapid detection of apoptosis in a manner that minimizes processing time and optimizes stable assay conditions. The purpose of this paper is to describe an automated method for the detection of apoptosis using the principle of microcapillary action for reagent uptake in combination with enzymatic nucleotidyl end-labeling of apoptosis-induced fragments of DNA and immunohistochemical detection.

Materials and Methods

Instrumentation and Pre-Run Setup

The detection of digoxigenin-labeled genomic DNA using direct immunoperoxidase was performed using the ApopTag™ *in situ* apoptosis detection kit/peroxidase (ONCOR, Gathersberg, MD) and the TechMate 1000™ flexible staining system (BioTek Solutions Inc, Santa Barbara, CA). This automated system uses the principle of capillary gap action to deliver and retrieve reagents used to process tissue specimens attached to positively charged glass slides (eloquently illustrated by Iezzoni *et al.*, 1993). The system is also equipped with an oven chamber capable of incubating the slides at 37° C while maintaining the appropriate humidity.

In order to reduce the amount of reagent needed for each set of slides, care was taken to minimize the area on the bottom of the well that was occupied by the reagent. Using a pipette, the reagents were placed in the center of the well bridging the side walls of the well. Minimizing the surface area covered by reagent allowed surface tension to maintain maximum height of the droplet of reagent, ensuring contact of

reagent with the bottom edge of the slides during placement of the slide set by the TechMate robotic arm for initiation of microcapillary action.

Tissue Processing

This protocol was optimized in tissues known to undergo apoptosis. Four different tissue types (liver, mammary gland, uterus and embryonic limb bud), all exhibiting either physiologic or chemically-induced apoptosis, were chosen to demonstrate the labeling of apoptotic cells. Specifically, hepatocellular apoptosis was induced in a B6C3F1 male mouse exposed by inhalation to 2.50 mg chloropentafluorobenzene (CTFB) per liter for 6 hours per day for 15 days. The uterine specimen was taken from an untreated female rat during endometrial mucosal and glandular involution associated with her normal reproductive cycle. Animals were euthanized using CO_2 or halothane and the respective organs removed and immersion-fixed in 10% neutral-buffered formalin for an undetermined amount of time, but at least 72 hours. Mammary gland undergoing post-lactational involution was excised from a normal female rat 3-5 days after weaning of pups and a paraffin block of this tissue was purchased from ONCOR (Gathersberg, MD) for use as control tissue with the ApopTag™ kit. The embryonic hindlimb bud (gestation day 11) was obtained by removing a whole embryo from a CD-1 mouse (Charles River Breeding Laboratories, Portage, MI) one hour after the dam was given a single oral gavage of all-*trans* retinoic acid in soybean oil (100 mg/kg body weight) and then euthanized by CO_2 asphyxiation. The whole embryo was immediately placed in 10% buffered formalin for 18 hours at 4°C, then transferred to 70% ethanol at the same temperature until dehydrated, cleared, and embedded in paraffin wax.

All specimens were cut four µm-thick, mounted on positively charged ChemMate slides (BioTek Solutions Inc , Santa Barbara, CA) and air or oven dried. Immediately before beginning the automated procedure, the slides were deparaffinized and hydrated in distilled water.

Specimen Preparation for Automated In situ Detection

The placement of the tissue on the glass slides and alignment of the slides in our procedure was modified from Iezzoni *et al.* (13). To reduce the amount of costly reagents and optimize microcapillary action, we further refined this procedure by placing tissue sections on the lower 1/4 to 1/3 of the slide, with placement slightly to the right of the center of the slide. For example, two specimens placed slightly off center to the right on different slides, when placed together facing each other, would not overlap. This improved the wicking action and staining homogenicity in sample pairs. Placing the tissue sections too far toward the top of the slide will require the use of more reagent to assure that the tissue specimen is completely covered by reagents.

Before beginning the deparaffinization procedure, slide surfaces were carefully cleaned to remove any dust particles attached to the positively charged slides. Following deparaffinization, slides were rinsed

in distilled/deionized water. Latex or nitrile gloves were always worn when handling slides to prevent deposit of natural oils in the skin onto the surfaces of the glass slides which could potentially deminish the efficacy of the microcapillary action. Slides were grouped in pairs, and each pair primed by completely submerging the pair in a Copland jar containing buffer A, swirling them about to insure complete wetting of both slides, and carefully bringing the slides together (specimens facing each other), while submerged. The slides have spacers that form the precision microcapillary gap and prevent any part of the tissue specimens from coming into contact should they overlap when the slide pair is loaded into the cassette. The slides were checked for a complete layer of buffer A between them. Carefully, the slides were aligned while still submerged, removed from the buffer and the pair placed in the cassette rack. After all slide pairs had been placed in the cassette rack, the final realignment of the slide edges was done using a flat side of a polypropylene five-histoslide transport container (Fisher Scientific, Pittsburgh, PA). This alignment made certain that all slides came into contact with the reagents and absorptive pads at the same time enhancing microcapillary action. To check for proper wicking action, the cassette was drained and reloaded with buffer A several times using an extra reagent tray and an absorbent pad. If filling or draining of slide pairs was incomplete, the realignment process was repeated.

***In situ* Detection**

An automated 76 step procedure was developed to perform the *in situ* detection (Table 1). The arrangement of the reagents used at each of the workstations is shown in Figure 1. The composition and volumes of the reagents, along with some commercially available sources for supplies are listed in Table 2. A brief description of the procedure follows, but for exact times of steps and amounts of reagents, refer to Tables 1 and 2.

The cassette containing the slide pairs primed with buffer A was positioned at home 1 work station (Figure 1) and the automated procedure started. After draining buffer A on pad 1 (Figure 1), the slides were incubated with a diluted trypsin solution (1:250, Boehringer Mannheim, Indianapolis, IN) at 37°C for 15 minutes. Slides were then rinsed in 4 changes of distilled water and endogenous peroxidase was quenched by exposing the tissue to 3% hydrogen peroxide in dH_2O for 6 minutes. Following exposure to an equilibration buffer (ONCOR, S7100-1, Gathersberg, MD) for 1 minute, sections were incubated at 37°C for 1 hour with working-strength terminal deoxynucleotidyl transferase (ONCOR, S7100-3, Gathersberg, MD) and reaction buffer (ONCOR, S7100-2, Gathersberg, MD) at a ratio of 1:2.38, respectively. The terminal deoxynucleotidyl transferase is used to catalyze the reaction by which digoxigenin-nucleotides are incorporated into the 3'-OH ends of double- or single-stranded DNA breaks. Following this incubation, slides were again incubated with a working-strength stop/wash buffer (ONCOR, S7100-4, Gathersberg, MD) at 37°C for thirty minutes. A peroxidase-linked anti-digoxigenin antibody with a removed Fc region

(ONCOR, S7100-5, Gathersberg, MD) was then applied to tissue sections. This allowed for optimal labeling of new 3'-OH double-stranded DNA breaks that are known to occur in high concentrations during apoptosis. After 3 rinses in H_2O, the diaminobenzidine (DAB) chromagen was added to the sections for 3 minutes. The DAB is known to react with the bound peroxidase that is present on the anti-digoxigenin antibody thus enabling localization of the 3'-OH double stranded DNA breaks. Tissue sections were counterstained with hematoxylin (BioTek Solutions Inc, Santa Barbara, CA) for 1 minute, dehydrated in xylene and coverslips were mounted using Permount™.

Complete removal of any residual reagent between the slides before critical steps was essential to maintaining the correct volumes and concentrations of reagents important for consistency of results. Critical steps (denoted by * in Table 1) were identified in the procedure, and slide pairs were visually checked just before each of these steps for complete reagent removal by the absorptive pad. If removal of reagents was incomplete, the machine was stopped, the slide rack removed, slides aspirated manually using a vacuum apparatus (Figure 2), the slide rack reattached to the robotic arm, and the automated protocol continued.

Results

Careful placement of the reagents into the center of the wells of the polypropylene trays, allowed a reduction of 25 percent of the volume of critical reagents (see reagents marked * in Table 1) required per specimen compared to the manual method.

Tissues processed and stained with the automated system described above are shown in figures 3 through 6. The processing time required was approximately four hours. In the specimens studied, 85 - 95% of the cells or cellular fragments that were morphologically consistent with apoptotic cells or bodies (solid arrows) stained golden-brown immunohistochemically with the chromogen, diaminobenzidine. The apoptotic cells usually occurred singly or in aggregates of only a few cells. Many of the apoptotic cells were shrunken and had lost contact with adjacent cells. Nuclear fragmentation (karyorrhexis) and condensation of chromatin (pyknosis) was evident. These nuclear remnants frequently were packaged in a narrow rim of membrane-delimited cytoplasm, collectively forming apoptotic bodies. Phagocytosis of these bodies by neighboring parenchymal cells was often present, and was particularly evident in the epithelial cells of the endometrium (Figure 3, arrow heads). The absence of inflammatory cells in all the specimens was a hallmark, distinguishing apoptosis from necrosis (cell death caused by exogenous insult). Cells undergoing mitosis, which were frequent in the embryonic limb bud, did not stain immunohistochemically (Figure 6, hollow arrows).

Discussion

Close attention to specimen preparation and instrument setup before the assay run, as well as intervention with procedural modifications when necessary in association with critical steps added reproducibility to the assay.

Our system has a built-in *in situ* oven that incubates the slides in a special rack at a programmable preset temperature. This oven feature holds the temperature and humidity constant while precisely timing the incubation period. Not all automated systems have the *in situ* oven feature, in that situation, the cassette can be removed from the robotic arm and placed in a 37°C incubation oven (humidity 98%) for the required time. At the end of the incubation period, the cassette can be placed back on the robotic arm, and the automated procedure continued.

A major advantage of this procedure is the ability to use small amounts of reagents reducing the expense of processing large numbers of slides. For example, the working concentration of terminal deoxynucleotidyl transferase (TdT), the amount recommended for each slide according to the directions in the ApopTag kit is 54μl or 108μl for two slides. Using this automated system allowed the use of 81μl of TdT, a reduction of 25% of reagent.

For reagents where limited volumes are used, it is best to load the wells 15 minutes prior to the step. The surface tension on the polypropylene trays is optimum up to approximately 15 minutes. This insures maximum height of the reagent droplet for contact with the lower margin of the microcapillary slides. Alternatively, a region of the well can be outlined using a PAP-PEN® (The Binding Site, San Diego, CA). An effective way to block off a small region in the center of the well is to outline a small square in the bottom of the well using a tooth pick dipped in the water repellent barrier liquid contained in the pen. This forms a barrier retaining the reagent liquid in a small area and prevents it from spreading over the entire well.

The placement of the tissue section onto the glass slide is critical in reducing the amount of reagent used. Ideally, tissue sections should be placed on the lower 1/4 of the slide. The amounts of reagents listed should be enough to cover specimens which are approximately 0.5 - 1 cm^2. If the specimens also are placed slightly off center to the right on both slides, the specimens are likely not to overlap when the slides are paired. This significantly improves the uptake and removal of reagent, producing stained specimens of superior quality.

Slide alignment in the cassette is critical to allow optimum microcapillary action and wicking of reagents. Use of a flat, broad surface such as that provided by a rectangular polypropylene histoslide transport box placed across the bottom edges of the paired slides is ideal. The polypropylene is soft and prevents chipping of the bottom end surfaces of the glass slides. It is these surfaces that make contact with the reagents during reagent uptake. Chipping reduces the effectiveness of reagent uptake and removal.

Occasionally, some slide pairs will not drain properly, leaving residual reagent before entering into a critical step. This can dilute the concentration of the critical reagent, thereby interfering with the optimum

assay conditions. By using the aspiration device (Figure 2) before critical steps in the assay, any residual reagent left between the slides can be removed allowing the appropriate amount of the critical reagent, thereby improving uniformity of assay results.

Numerous wash or rinse steps intervene between critical steps in the protocol. We have used both dH_2O and phosphate-buffered saline (PBS) for wash or rinse steps, and found that either works equally well.

There are several to using the automated system. One advantage of using paired slides during the assay procedure is that it allows the investigator to optimize the experimental design. Each specimen per slide receives exactly the same treatment during processing. For example, a specimen obtained from a high dose animal could be paired with a control specimen. This assures that both specimens receive exactly the same processing adding credibility to any differences detected between control and treatment samples. Moreover, the precise timing of each step by the automated system and the processing of up to 40 slides (10 pairs) at one time, using the same freshly prepared batches of reagents, help standardize and make this procedure very reproducible. Simultaneous batch processing of a relatively large number of samples simultaneously has the added advantage of being able to include equal numbers of control and one or more treatment specimens per automated run. For example, a maximum of 40 slides per run can be comfortable managed. That allows 10 specimens each from control, low, medium and high exposure groups. If more than one automated run is required to complete all the specimens, any between run variability due to processing is equally distributed across all experimental groups. This essentially normalizes for the inter-run variablilty in a study. When doing this assay manually, it is virtually impossible to handle 20-40 specimens, timing each step exactly throughout the lengthy procedure. It also is evident that the use of the automated system significantly reduces the technician's time, allowing for more productivity in the laboratory.

A slight limitation of the procedure was that approximately five to fifteen percent of the cells morphologically consistent with apoptotic cells or apoptotic bodies did not appear to label using the automated system with the ApopTag™ kit. Residual unlabeled cells were most likely due to less than optimum DNA retrieval by enzyme digestion of tissue.

Conclusions

We believe that the use of this automated method of *in situ* detection of apoptosis is a potentially valuable research tool which can significantly imporve the reproducibility of experimental results, conserve the technologist's time and effort, and reduce the quantity of reagents required. This equates to improved quality control and lower research costs. Additionally, many of the procedural steps described also can be applied to more conventional automated immunohistochemical methods; particularly when detection and relative quantification of nonisotopic, colorimetric signal using computer-based image analysis is important.

Acknowledgements

This research was supported by the Strategic Environmental Research and Development Program and the US Air Force Office of Scientific Research Summer Research Program for Faculty to MAS. The animals used in this study were handled in accordance with the principles stated in the Guide for the Care and Use of Laboratory Animals, prepared by the Committee on Care and Use of Laboratory Animals of the Institute of Laboratory Animals Resources, National Research Council, Department of Health and Human Services, National Institute of Health Publication No. 86-23, 1985, and the Animal Welfare Act of 1966, as amended. The authors would like to thank all the personnel from the pathology branch for their technical assistance, USAF Lt. Steve Nystrom for his professional art work, and Ms. Susan Godfrey for her editorial assistance.

References

1. Clark, PGH: Developmental cell death: morphological diversity and multiple mechanisms. *Anat Embryol* 181:195-213, 1990.

2. Compton MM: Thymocyte apoptosis - a model of programmed cell death. *Trends Endo* 3:17-23, 1992.

3. Kerr JF, Wyllie AH, Currie AR: Apoptosis: a basic biological phenomenon with wide ranging implications in tissue kinetics. *Brit J Cancer* 26:239-257, 1972.

4. Bursch W, Paffe S, Pulz B, et al: Determination of the length of the histological stages of apoptosis in normal and in altered hepatic foci of rats. *Carcinogenesis* 11:847-853, 1990.

5. Allen DJ, Roberts SA, Harmon BV: Spontaneous spermatogonial apoptosis shows three distinct morphological phases and no circadian rhythm in the rat. *Cell Proliferation* 25:241-250, 1993.

6. Kyprianou N, Isaacs J: Activation of programmed cell death in the rat ventral prostate after castration. *Endocr* 122:552-562, 1988.

7. Nakamura T, Sakai T, Hotchi M: Histochemical demonstration of DNA double strand breaks by in situ 8'-tailing reaction in apoptotic endometrium. *Biotech Histochem* 70:33-39, 1995.

8. Alison MR, Sarrof CE: Apoptosis: regulation and relevance to toxicology. *Hum Exp Toxicol* 14:234-247, 1995.

9. Williams GT: Programmed cell death: apoptosis and oncogenesis. *Cell* 65:1097-1098, 1991.

10. Goldsworthy TL, Frannson-Steen R, Maronpot RR: Importance of and approaches to quantification of hepatocyte apoptosis. *Toxicol Pathol* 24:24-35, 1996.

11. Alison MR, Sarraf CE: Apoptosis: a gene-directed programme of cell death. *J R Coll Physicians Lond* 26:25-35, 1992.

12. Vaux DL: Toward an understanding of the molecular mechanisms of physiological cell death. *Proc Natl Acad Sci* 90:786-789, 1993.

13. Iezzoni JC, Manahan LJ, Park C-S et al: Recent advances in automated immunocytochemistry. *J Histotechnol* 16:39-50, 1993.

TABLE 1: Program for Automated Detection of Apoptosis Using the ApopTag™ *In Situ* Detection Kit

Step	Event	Time (Hr:Min:Sec)	Temperature[1] (°C)
1	Drain on Pad #1	00:00:29	
2	Apply trypsin	00:00:15	
3	Incubate with trypsin in humidified chamber	00:15:00	37
4	Drain on Pad #1	00:00:29	
5	Rinse with dH_2O solution	00:02:00	
6	Drain on Pad #1	00:00:29	
7	Rinse with dH_2O solution	00:02:00	
8	Drain on Pad # 1	00:00:29	
9	Rinse with dH_2O solution	00:02:00	
10	Drain on Pad #1	00:00:29	
11	Rinse with dH_2O solution	00:02:00	
12	Drain on Pad # 1	00:00:45	
13	Rinse with H_2O_2 solution	00:02:00	
14	Drain on Pad # 1	00:00:29	
15	Rinse with H_2O_2 solution	00:02:00	
16	Drain on Pad # 1	00:00:29	--
17	Rinse with H_2O_2 solution	00:02:00	
18	Drain on Pad # 1	00:00:45	
19	Rinse with Buffer A	00:00:10	
20	Drain on Pad # 1	00:00:29	
21	Rinse with Buffer #A	00:00:10	
22	Drain on Pad # 1	00:00:29	
23	Rinse with Buffer A	00:00:10	
24	Drain on Pad # 1	00:00:45	
25	Rinse with Buffer B_1	00:01:00	

26	Drain on Pad # 2	00:00:29	
27	Rinse with Buffer B_1	00:01:00	
28	Drain on Pad # 2	00:00:29	
29	Rinse with Equilibration Buffer	00:01:00	
30*	Drain on Pad # 2	00:00:45	
31	Apply TdT	00:00:20	
32	Incubate with TdT in humidified chamber	1:00:00	37
33*	Drain on Pad # 2	00:00:45	
34	Rinse with Stop/Wash	00:00:20	
35	Incubate in humidified chamber	00:30:00	37
36	Drain on Pad # 2	00:00:29	
37	Rinse with Buffer B_1	00:02:00	
38	Drain on Pad # 2	00:00:29	
39	Rinse with Buffer B_1	00:00:10	
40	Drain on Pad # 2	00:00:29	
41	Rinse with Buffer B_1	00:00:10	
42	Drain on Pad # 2	00:00:29	
43	Rinse with Buffer B_1	00:00:10	
44*	Drain on Pad # 2	00:00:45	
45	Apply Anti-DAB	00:30:00	
46	Drain on Pad # 3	00:00:29	
47	Rinse with Buffer B_1	00:00:10	
48	Drain on Pad # 3	00:00:29	
49	Rinse with Buffer B_1	00:00:10	
50	Drain on Pad # 3	00:00:29	
51	Rinse with Buffer B_1	00:00:10	
52	Drain on Pad # 3	00:00:29	
53	Rinse with Buffer B_1	00:00:10	
54*	Drain on Pad # 3	00:00:45	
55	Apply DAB	00:03:00	

56	Drain on Pad # 3	00:00:29	
57	Rinse with Buffer B_2	00:00:10	
58	Drain on Pad # 3	00:00:29	
59	Rinse with Buffer B_2	00:00:10	
60	Drain on Pad # 3	00:00:29	
61	Rinse with Buffer B_2	00:00:10	
62	Drain on Pad # 4	00:00:29	
63	Rinse with Buffer B_2	00:00:10	
64	Drain on Pad # 4	00:00:29	
65	Rinse with Buffer B_2	00:00:10	
66	Drain on Pad # 4	00:00:45	
67	Apply Hematoxylin stain	00:01:00	
68	Drain on Pad # 4	00:00:29	
69	Rinse with dH_2O solution #2	00:00:10	
70	Drain on Pad # 4	00:00:29	
71	Rinse with dH_2O solution #2	00:00:10	
72	Drain on Pad # 4	00:00:29	
73	Rinse with dH_2O solution #2	00:00:10	
74	Drain on Pad # 4	00:00:29	
75	Rinse with dH_2O solution #2	00:00:10	
76	Home Station	-------	

[1]Conducted at room temperature unless otherwise indicated.

Figure 1. Diagram of Work Station Layout.

	1	2	3	4	5	6	
D	Home 1	Buffer A	Buffer B_1	Buffer B_2	Equilibration Buffer		
C	Home 2	Pad 1	Pad 2	Pad 3	Pad 4		Oven
B	Home 3	Stop Wash	Hydrogen Peroxide	1 H_2O	2 H_2O	DAB	
A	Home 4	Trypsin		TdT	Anti-DAB	Hematoxylin	

Table 2. Composition of Reagents and Other Supplies

I. ApopTag™ Kit

A. Terminal deoxynucleotidyl transferase (TdT):
   a. 57 µl per well of Reaction buffer (provided in kit)
   b. 24 µl per well of TdT concentration stock (provided in kit)
   c. Vortex, prepare fresh and store on ice not more than 6 hrs
   d. Add 81µl of working stock per well

B. DAB substrate:
   a. 117 µl DAB Dilution buffer per well
   b. 13 µl DAB chromagen per well
   c. Vortex, prepare fresh
   d. Add 130 µl DAB solution per well

C. Stop/Wash Buffer:

   a. 1 ml concentrated Stop/Wash buffer (S7101-4)

   b. 34 ml dH_2O
   c. Vortex, working solution can be stored for up to 1 year in a glass or plastic container at 4°C.
   d. Add 300 µl of working solution per well

D. Equilibration Buffer (working solution provided in kit): use 107 µl per well

E. Antibody (working solution provided in kit): use 81 µl per well.

II. Other Reagents

A. Hydrogen Peroxide (H_2O_2):
  a. Make up a 1:10 dilution of 30% H_2O_2 solution:dH_2O
  b. Add 0.15% (0.15 ml/100 ml) Tween 20® (Calbiochem-Novabiochem, La Jolla, CA)
  c. Prepare fresh
  d. Add 1 ml per well

B. Trypsin - 1:250; use 300 μl per well (Boehringer Mannheim, Indianapolis, IN)

C. Buffer A (1000 ml, pH 7.2)
  a. 100 ml 10x PBS Buffer (Boehringer Mannheim, Indianapolis, IN, Stock # 100 961)
  b. 1.5 ml of Tween 20® (Calbiochem-Novabiochem, La Jolla, CA)
  c. 500 mg Sodium Azide (Sigma, St. Louis MO, Stock # S8032)
  d. 650 g Bovine Serum Albumin (Sigma, St. Louis, MO, Stock # A2153)
  e. Adjust pH and QS to 1000 ml with dH_2O (Shelf life at least 6 months)
  f. Add 1 ml per well

D. Buffer B (1000 ml, pH 7.2):
  a. 100 ml 10x PBS Buffer (Boehringer Mannheim, Indianapolis, IN, Stock # 100 961)
  b. 1.5 ml of Tween 20® (Calbiochem-Novabiochem, La Jolla, CA)
  c. 100 mg Thimersol (Sigma, St. Louis MO, Stock # T8784)
  d. 100 mg Gentamicin Solution (Gibco BRL, Life Technologies, Inc., Grand Island, NY, Stock # 15750-011)
  e. Adjust pH and QS to 1000 ml with dH_2O (Shelf life at least 6 months)
  f. Add 1 ml per well

E. dH_2O:
  a. Add 0.15% (1.5 Fl/ml) Tween 20® (Calbiochem-Novabiochem, La Jolla, CA).
  b. Vortex, use 1 ml per well

F. Hematoxylin stain: Add 300 μl per well (BioGenex, Stock #HK 100-5K, San Ramon, CA).

III. Reagent Wicking (Absorptive Pads): (Curtin Matheson Scientific, Stock # 314-259, Cincinnati, OH or BioGenex, Stock # XT007-WP, San Ramon, CA

IV. Reagent Trays (BioGenex, Stock # XT008-2T, San Ramon, CA

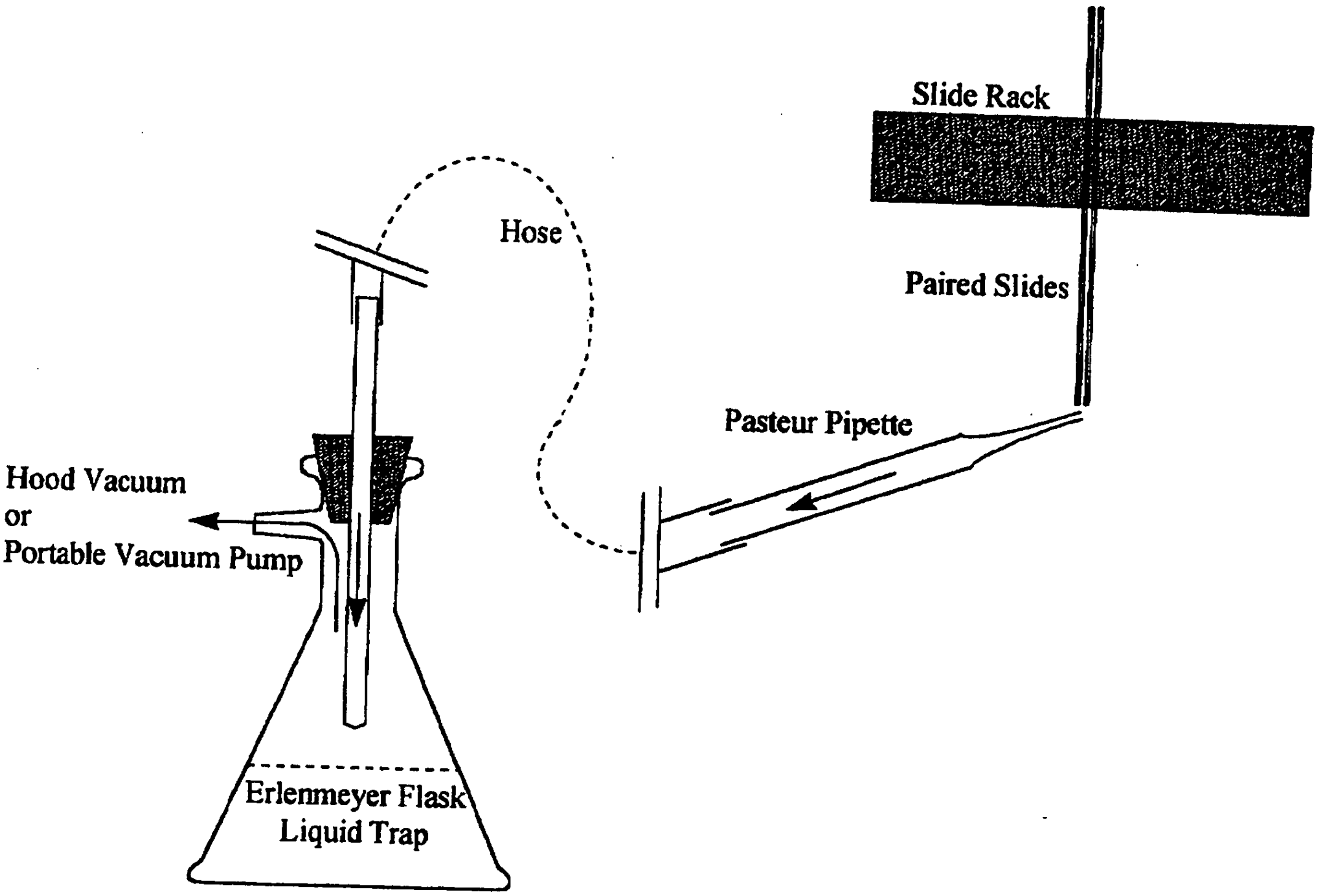


Figure 2. Vacuum Apparatus. The aspiration apparatus consisted of a Pasteur pipette attached to a vacuum hose with a typical in-line Erlenmeyer flask liquid trap and with vacuum supplied by a hood vacuum or a portable vacuum pump.

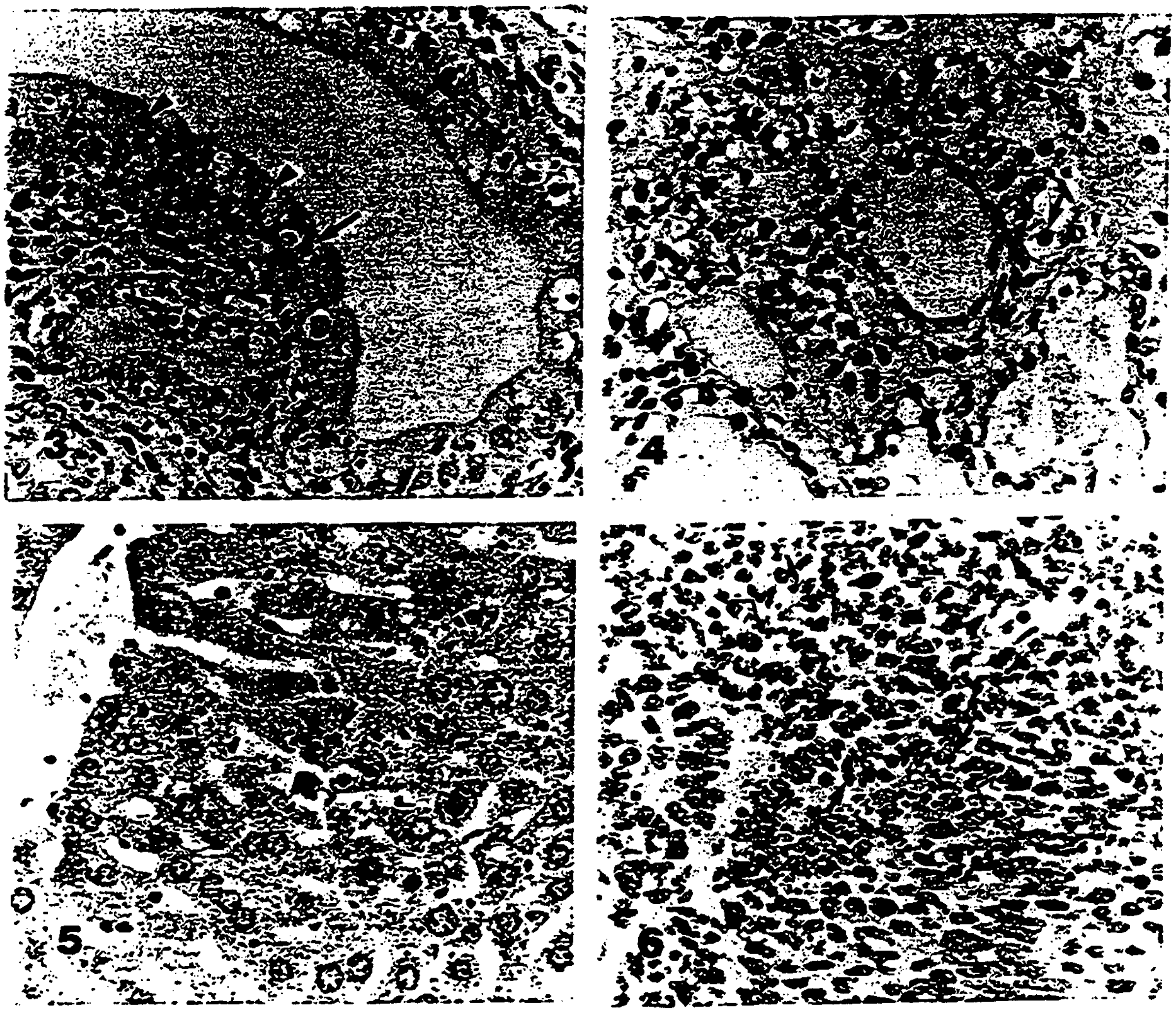


Figure 3. Photomicrograph of rat uterus. The golden-brown foci (solid arrows) are apoptotic epithelial cells or fragments (apoptotic bodies) of the endometrium and subjacent endometrial glands. Note the phagocytized apoptotic bodies in the cytoplasm of epithelial cells of the endometrium (arrow heads). Stained immunohistochemically with a hematoxylin counterstain. x500

Figure 4. Photomicrograph of rat mammary gland undergoing post-weaning involution. Arrows point to apoptotic secretory epithelial cells. Stained immunohistochemically with a hematoxylin counterstain. x500.

Figure 5. Photomicrograph of mouse liver manifesting a small aggregate of apoptotic bodies (arrow) derived from one or two hepatocytes. Apoptosis was induced by inhalation exposure to chloropentafluorobenzene (CPFB). Stained immunohistochemically with a hematoxylin counterstain. x500.

Figure 6. Photomicrograph of a mouse embryonic limb bud illustrating numerous apoptotic mesenchymal cells (arrows) induced by the administration of an oral dose of all-*trans* retinoic acid. Note that cells in mitosis (hollow arrows) do not stain (as they should not). Stained immunohistochemically with a hematoxylin counterstain. x500.

CONCLUSIONS

Objective number 1 was met with the drafting and submission of the manuscript found in the previous pages. Objective number 2 is still underway, but the findings to date are given in the following abstract.

Apoptosis, or programmed cell death, occurs during embryonic development during normal differentiation or from toxic insult. Locating apoptosis in an embryo is difficult due to few morphological markers and the size of the embryo. This study identified regions of apoptosis in the limb bud (LB) when the embryo was susceptible to limb malformations from exposure to exogenous substances. Pregnant CD-1 mice were administered all-*trans* retinoic acid on gestation day 11 and embryos harvested at various times after administration. Whole embryos were placed in buffered formalin, dehydrated, and embedded in paraffin. Serial sagittal sections were processed and stained with either hematoxylin and eosin (H&E) or with ApopTag™. Areas of apoptosis were documented with photomicrographs. The marker most important in locating the maximum area of LB apoptosis was the central artery (CA) in both control and treated embryos. The area extended from the CA toward the trunk. Two other landmarks, the apical ectodermal ridge (AER) and nerve trunk extending into the LB, were helpful in indicating proper depth and longitudinal orientation. Sections lacking the CA had fewer or no apoptotic bodies although other areas of the same LB might show large areas of apoptosis. Treated animals had larger areas of apoptosis and more apoptotic bodies than control animals. Treated animals often had additional areas of apoptosis around the marginal vein. There was no visible difference between the location or amount of apoptosis using H&E or ApopTag™ staining.

BIOREMEDIATION AND ITS EFFECT ON TOXICITY

Daniel P. Smith
Department of Civil and Environmental Engineering
Utah State University
Logan, Utah 84322-8200

Final Report for:
Summer Faculty Research Program
Armstrong Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, DC

and

Armstrong Laboratory

August 1996

BIOREMEDIATION AND ITS EFFECT ON TOXICITY

Daniel P. Smith, Professor
Department of Civil and Environmental Engineering
Utah State University
Logan, Utah 84322-8200

Abstract

Bioremediation of petroleum hydrocarbons and its effect on toxicity reduction was investigated through a combination of literature and modeling studies. A literature review was conducted on the biotransformation of petroleum hydrocarbons and microbial metabolite formation under alternate electron acceptor conditions when oxygen is not present. The literature review provided the basis for developing a predictive model of biodegradation of petroleum hydrocarbons under anaerobic conditions. A multispecies energetic/kinetic model of alkylbenzene biodegradation was developed and applied to quantitatively predict toluene mineralization by anaerobic consortia when ferric iron and sulfate were available electron acceptors. Simulations predicted that iron was preferentially used as the electron acceptor when iron- and sulfate-reducing microorganisms were present at equal initial populations.

The chemical properties of specific components of JP-4 and potential microbial degradation products were modeled using structure-based Group Contribution Methods. Metabolite formation from parent compounds generally enhanced water solubility and mobility, suggesting that the concentrations and transport properties of metabolites may be important factors in toxicity and risk reduction assessments. A literature review was conducted on the effects of bioremediation on the reduction of toxicity of soils and groundwaters contaminated with petroleum hydrocarbons. The literature results confirm that bioremediation generally results in a reduction in toxicity as measured by a variety of acute and chronic assays, and that toxicity reduction is often corroborated with the reduction in concentration of specific quantifiable chemical components. The results of this literature and modeling study provide a basis with which to develop follow-up laboratory, field, and modeling studies of bioremediation of petroleum hydrocarbons and its effect on toxicity and risk reduction.

BIOREMEDIATION AND ITS EFFECT ON TOXICITY

Daniel P. Smith, Professor
Department of Civil and Environmental Engineering
Utah State University
Logan, Utah 84322-8200

Introduction

The environmental impact of jet fuels and other petroleum products is of great concern to the United States Air Force. Understanding of the biological transformations, chemical properties that affect contaminant mobility and bioavailability, and the toxicity of parent jet fuel components and bioremediation metabolites is needed to support human health and ecological risk assessments for contaminated subsurface sites. This report presents the results of literature and modeling investigations into biological transformations of hydrocarbons in the subsurface and their potential toxicity to human and ecological receptors.

The objectives of this work were to perform a literature review on biological transformations of petroleum hydrocarbon components when released into the subsurface, to develop structured biochemical models that could quantitatively predict biodegradation of hydrocarbon components and formation of metabolites, to examine the chemical properties of jet fuel components and metabolites that affect subsurface fate and transport, and to review the effects of bioremediation on toxicity of soil and groundwater. A literature review was conducted to evaluate biological transformations and metabolite formation under anaerobic conditions. A body of references was collected, reviewed, and searched for relevant technical content and reference sources. A computerized search of numerous databases was completed. Further references cited in these articles were then retrieved. The databases searched included NTIS, Compendex, Geoarchive, Energyline, Georef, Ei Meetings, P/E News, Chemical Abstracts, and Water Resources. The combined papers were organized by topic, and those papers addressing the same topic were simultaneously reviewed. Using a similar approach, a literature search was conducted on the effect of bioremediation processes on changes in toxicity at subsurface sites contaminated with petroleum products. Anaerobic biological transformations of simple alkylbenzenes were simulated by structured, multispecies biochemical modeling and numerical integration techniques. The estimation of chemical properties of jet fuel components and metabolites was performed using Group Contribution Methods contained in the DESOC (171).

This report is based on a longer report titled *Subsurface Bioremediation of Hydrocarbons and Its Effect on Toxicity*, submitted to Armstrong Laboratory in August 1996.

Anaerobic Transformations of Petroleum Hydrocarbons

Though petroleum hydrocarbons biodegrade most readily when with oxygen supplied either naturally or by engineered processes (2,4,208), intrinsic remediation most often relies on the presence of alternate electron acceptors (4). When oxygen is depleted, the common electron acceptors include nitrate, iron, and sulfate, as well as

inorganic carbon in methanogenic processes. In addition, humic substances, which are ubiquitous in natural environments and are formed from microbial processes (214,215), may serve as microbial electron acceptors (216). This section reviews studies that have demonstrated biotransformation of petroleum hydrocarbons under denitrifying, sulfate reducing, and methanogenic processes (28,44,52,53,90-96,100,101,135,141).

Many organic compounds associated with petroleum product usage are potentially biodegradable in intrinsic subsurface processes (88, 140, 142). Transformations are effected by the structure of the organic molecule, the concentration of the compound and possible toxicity of the compound, its degradation products, or other organic or inorganic materials, the electron acceptors present, pH, temperature, moisture content, dissolved solids, macro- and micro-nutrients, the bioavailability of the compound to attached or suspended microorganisms as influenced by sorption or the presence of nonaqueous phase liquids, and the presence of organics which may serve as primary substrates for cometabolic reactions or competitively inhibit a target compound biodegradation (88, 130).

An energy balance model for microbial metabolism was used to generate overall stoichiometries including microbial synthesis for complete mineralization of toluene under denitrifying, sulfate reducing, and methanogenic terminal electron acceptor processes (94, 98). These reactions are shown in Table 1. The stoichiometric relationship between toluene consumption and electron acceptor requirement (NO3 and SO4) or methane production are summarized in Table 2. The stoichiometric values for (fs)max represent the maximum amount of electron donor used for synthesis (no endogenous respiration), while those for(fs)min apply where substantial cell decay has taken place. Observed stoichiometries should therefore fall somewhere between these two extremes. The variation in stoichiometry between (fs)max and (fs)min is greater for anaerobic processes that yield more free energy per electron transferred, such as denitrification, than for lower energy yielding processes. The lower energy yielding processes such as methanogenesis have overall stoichiometries that are closer to the catabolic reaction because of the small fraction of electron equivalents used for bacterial synthesis even under maximum synthesis conditions. Biochemical reactions have been derived for two to four ring polyaromatic hydrocarbons using electron acceptors which are potentially available in the subsurface (28).

The terminal electron acceptor is one the most significant practical influences ~~on~~ intrinsic bioremediation. Hydrocarbon plumes show different zones which are delineated based on the predominant electron acceptor present. An example is the establishment of sequential methanogenic/sulfidogenic and ferrogenic zones downgradient from

Table 1
Biochemical Stoichiometry for Toluene Mineralization Under Electron Acceptor Conditions
(f_s = 0.60 (f_s)max)

Denitrification

$$C_7H_8 + 5.602\ NO_3^- + 0.338\ H^+ = 0.347\ C_5H_7O_2N + 2.628\ N_2 + 5.263\ HCO_3^- + 0.324\ H_2O$$

Sulfate Reduction

$$C_7H_8 + 4.262\ SO_4^{-2} + 0.0950\ NH_4^+ + 1.912\ H^+ + 2.174\ H_2O = 0.0950\ C_5H_7O_2N + 4.262\ H_2S + 6.525\ HCO_3^-$$

Methanogenesis

$$C_7H_8 + 0.0583\ NH4^+ + 0.0583\ HCO_3^- + 4.767\ H_2O = 0.0583\ C_5H_7O_2N + 4.356\ CH_4 + 2.412\ CO_2$$

Table 2
Stoichiometric Relationships for Anaerobic Toluene Biodegradation

Terminal Electron Acceptor		$(f_s)_{max}$	$(f_s)_{min}$
Denitrification			
	NO3-N (g/g)	0.69	1.02
	VSS (g/g)	0.71	0.142
	HCO3 (eq/g)	+ 0.045	+ 0.070
Sulfate Reduction			
	SO4-S (g/g)	1.43	1.54
	VSS (g/g)	0.20	0.039
Methanogenesis			
	CH4 (g/g)	0.74	0.78
	VSS (g/g)	0.12	0.024

a municipal landfill (32) and also sequential zones of anaerobic (manganese and iron reduction and methanogenesis), low oxygen, and high oxygen conditions downgradient in a plume resulting from a crude oil spill (85). Other studies have shown that the terminal electron acceptor process operating at a particular subsurface location can be dynamic in time and space (31). Delivery of sulfate by infiltration or lateral advection can transform a methanogenic zone into a sulfidogenic zone, and the shift back to methanogenesis can occur if sulfate is depleted due to lack of inflow (66).

Terminal electron acceptor shifts occurred in time scales of 10 days to three months (31). Slow mixing rates of aquifer water can result in steep chemical gradients, and substantial differences in chemical parameters such as dissolved oxygen, over spatial scales of ten meters or less (42). Clayey, confining bed sediments often do not exhibit terminal electron accepting processes, but may store large quantities of sulfate and serve as an electron acceptor source to adjacent sand aquifers (43, 155). Sulfate transport from confining beds has been shown to support microbially mediated transformations of sedimentary organic matter in adjacent aquifers, thus alleviating apparent electron acceptor debt in the aquifer (43).

It is often difficult to establish the terminal electron acceptor processes which are operative in a given locality on the subsurface. Measurements of dissolved species usually do not show redox equilibrium, and redox electrodes respond to few of the significant redox couples in geochemical milieu (132). One potentially useful new technique is the use of dissolved molecular hydrogen concentrations to delineate the terminal electron acceptor processes operating in a subsurface locale (23,31,43,126,201). Reported ranges for molecular hydrogen levels are shown in Table 3 for denitrification, iron reduction, sulfate reduction, and methanogenesis. The dissolved hydrogen technique has the advantage that it theoretically measures an actual metabolic intermediate in an anaerobic terminal electron acceptor process, and it may have a future application in monitoring and characterizing intrinsic bioremediation processes in the subsurface. Dissolved hydrogen concentrations vary inversely with the amount of free energy available to the biochemical processes mediated under the different terminal electron acceptor processes. Energetic and kinetic analysis of electron accepting processes can be used to elucidate the nature of competition for molecular hydrogen and other anaerobic intermediates (10,185,187,196).

Table 3
Reported Dissolved Hydrogen Concentrations
in Sediment Pore Water (23, 43, 126)

Terminal Electron Acceptor	H_2 (nM)
Denitrification	0.05
Iron Reduction	0.1 -1
Sulfate Reduction	1 - 6
Methanogenesis	7 - 20

Denitrifying Conditions In laboratory batch microcosms of soils contaminated from a JP-4 fuel spill, rates of carbon mineralization and denitrification increased asymptotically to a maximum of 0.85 nmoles/g-hr at 1mM NO3, and were 38% lower at pH 4 than pH 7. CO2 production increased with added NO3, and N2 was the only product nitrogen species reported (45). Microcosm denitrification was NO3 limited and insignificant in the absence of added NO3; the rate of denitrification slowed with time, possibly reflecting the initial preferential consumption of easily oxidizable compounds at the expense of more persistent compounds. Denitrification rates in the batch microcosms increased from 0.05 to 0.72 nmole/g-hr as the total petroleum hydrocarbons increased from 26 to 390 mg/kg (45).

Toluene was completely transformed to CO2 and biomass in denitrifying cultures enriched from sediments, groundwater, contaminated soils, process effluent, and sludge, In all cultures, partial o-xylene degradation was observed and was dependent on toluene degradation (128). In another study, toluene was degraded by isolate T1, which was unable to grow on benzene, ethylbenzene, and xylenes. O-xylene was utilized only as a cometabolite with toluene, and unidentified intermediates accumulated from metabolism of toluene and o-xylene (47). These dead end metabolites were identified as benzylsuccinic acid and benzylfumaric acid from toluene degradation and (2-methyl-benzyl)succinic acid and (2-methyl-benzyl)-fumaric acid from m-xylene degradation (61). In another study of toluene degradation under denitrifying conditions, Pseudomonas strain K 172 mineralized toluene to carbon dioxide, with benzyl alcohol as an intermediate (143). This pure culture also used benzaldehyde and benzoate without a lag period.

Batch denitrification microcosm studies on aquifer cores from Traverse City, MI yielded biodegradation of toluene, ethylbenzene, o-xylene, m-xylene, and 1,2,4 trimethylbenzene, with no lag phase for toluene and 7 to 14 day lag periods for other degrading compounds (48). Benzene utilization was not observed, and o-xylene degradation ceased when the other degrading alkylbenzenes were depleted.

Toluene and m-xylene degradation were observed in a denitrifying laboratory column containing aquifer material from a river infiltration area, and numerous other alkylbenzenes and potential intermediates were found to be biodegradable under denitrifying conditions (49, 59, 125).

The degradation of napthalene and acenapthalene under denitrification conditions were investigated in laboratory microcosms containing previously uncontaminated soil, and found to proceed after 12 to 36 day acclimation times due to the need to increase the population of specific PAH degrading organisms. The PAH degradation rates were much slower than when natural soil organic carbon was used as electron acceptor (50).

Under nitrate limitation, preferential natural carbon utilization rendered the PAH stable. All of the PAH mass (aqueous + sorbed) was ultimately available for biological utilization. Previously uncontaminated soils were exposed to napthalene, and batch experiments were conducted to test for simultaneous solution phase microbial napthalene utilization and desorption. A radial diffusion model for napthalene sorption and desorption from soil particles was coupled with Michalis Menten utilization kinetics in the aqueous phase to predict the data. Though the soil acted as a reservoir for napthalene, prolonging the time required to deplete the aqueous phase, all of the napthalene was desorbed and degraded. The rate of napthalene utilization was slower than the rate of desorption, and equilibrium was established between sorbed and solution phase napthalene (51).

A field and laboratory study on Borden sand showed rapid toluene biotransformation under denitrifying conditions in both microcosms and in field tests (55). Here, ethylbenzene and xylene isomers were transformed to a lesser extent but benzene was not transformed in laboratory or field. Studies with suspended slurry reactors containing 30% solids indicated zero order kinetics for napthalene concentrations approaching the aqueous solubility and irreversible adsorption of napthalene which became more pronounced with higher soil organic content (58).

Laboratory treatability studies using previously uncontaminated aquifer material from Park City, Kansas were used to demonstrate utilization of toluene, m- and p-xylene, and trimethylbenzenes; benzene was not degraded and o-xylene degradation was appreciably less than the other xylene isomers (110).
Laboratory studies of BTEX compound degradation under denitrifying conditions in suspended growth and biofilm reactors have shown that toluene is preferentially utilized over ethylbenzene, that o-xylene cannot support growth of denitrifiers but is a cometabolite of toluene, that toluene can inhibit o-xylene cometabolism, and that benzene is not degraded in these reactor systems (111, 112).

Oxygenates added to gasoline have been shown to be biodegradable under denitrification conditions. Ethyl tertiary butyl ether (ETBE) and tertiary butyl alcohol (TBA) were degradable under denitrification conditions, with TBA having faster degradation rates (62). Degradation of both substrates was inhibited by addition of ethanol, which may have been preferentially utilized by denitrifiers. ETBE degradation was found only in soils with low organic carbon content. Methyl tertiary butyl ether (MTBE) was not biodegraded in the microcosms.

Sulfate Reducing Conditions There are reports that sulfate can be used an electron acceptor when hydrocarbons such as pentadecane, heptadecane, and octadecane are electron donors; sulfidogenesis decreased markedly on lower molecular weight hydrocarbons. Additionally, an anaerobic methane oxidation coupled to sulfate reduction has also been reported (96). Toluene was biodegraded in sulfate reducing conditions in soil microcosms and in enrichment cultures; toluene disappearance corresponded closely to sulfate disappearance (114). A possibly abiotic reaction between hydrogen sulfide and ferric iron resulted in the formation of reduced iron. Toluene carbon was over 80% mineralized to CO2 in enrichment cultures fed toluene and sulfate; < 10% of toluene carbon was converted to two dead end metabolites: benzylsuccinic acid and benzylfumaric (67). These same metabolites were also found in toluene utilizing, denitrifying culture of isolate T1 (61).

A natural gradient groundwater tracer test in a hydrocarbon contaminated aquifer showed degradation of TEX compounds, 1,3,5 TMB, and napthalene, but no benzene utilization; the first order rate constant for toluene degradation in a laboratory sand column was several hundred times higher than that calculated from field data where toluene concentration was much higher (68). o-Xylene showed more rapid degradation in the aquifer tracer test than did the other xylene isomers.

Sulfate reducing mixed cultures enriched from gasoline contaminated aquifer material on BTE, o-xylene, and p-xylene completely mineralized toluene and m -xylene; greater than 90% substrate carbon was detected as CO2 (69). Substrates were preferentially utilized in the order of toluene, p-xylene, and o-xylene. Benzene and ethylbenzene were not degraded. Sulfide inhibited the utilization rate of the monoaromatic hydrocarbons.

A sulfate reducing, hexadecane oxidizing organism was isolated from precipitates from an oil water separator (70). Hexadecane was completely mineralized to CO2 and cells. Strain Hxd3 could utilize C12 through C20 alkanes but not C < 12. Benzene was completely mineralized in sulfate reducing microcosms containing aquifer sediments from a contaminated subsurface site (76). Over 90% of benzene carbon was recovered as CO2, and SO4 was the presumed electron acceptor.

Methanogenesis Benzene and toluene were found to accumulate in methanogenic consortia enriched on ferulic acid, and benzene from benzoate, when methanogenesis was inhibited (71). Rather than being true metabolic intermediates, benzene and toluene may be formed as electron sink products when aceticlastic or hydrogenotrophic methanogens are suppressed. Toluene and o-xylene were degraded in methanogenic consortia enriched from gasoline contaminated aquifer material on toluene and/or 0-xylene (77). Toluene and o-xylene were transformed after 3 to 6 month acclimation periods, respectively, yielding 85 to 100% of theoretical methane. Transformation of these compounds was inhibited by other preferred substrates such as acetate, propionate, and H_2, suggesting that these or naturally occurring or co-contaminant organics may inhibit degradation of more difficult to degrade compounds.

Batch microcosms containing creosote contaminated aquifer material exhibited degradation of C3 to C6 aliphatic organic acids, accompanied by acetate accumulation and methanogenesis; temporal substrate patterns in the microcosm were observed in the downgradient direction in the aquifer (78).

Toluene and benzene were converted partially to methane in ferulate acclimated methanogenic consortia (115, 144). Significant intermediates in toluene degradation were o-cresol, p-cresol, and benzoic acid, and in benzene transformation were phenol, cyclohexanone, and propionic acid. Benzene and toluene transformations were considered to be probable fermentation reactions with the following possible initiation reactions: single ring hydroxylations, methyl oxidation of toluene, demethylation of toluene, and ring reduction. The finding that oxygen from water is incorporated into toluene and benzene transformations in methanogenic systems lends further support to the fermentative nature of these transformations (136).

Benzene was biodegraded in a submerged reactor packed with municipal solid waste. The mesophilic reactor supported acidogenic fermentation conditions with total volatile acids of 2100 to 4200 mg/l. Benzene declined from an initial concentration of 180 mg/l to nondetectable concentrations in 46 days (137).

Greater than 99% removal of toluene, ethylbenzene, benzene, and o-xylene were achieved after 120 weeks incubation in batch methanogenic microcosms containing alluvial aquifer material taken from a location adjacent to a municipal landfill in Norman, Oklahoma (116). Toluene was substantially degraded to CO2.

Anaerobic production and transformation of aromatic hydrocarbons was examined in a ferulate degrading, BESA inhibiting methanogenic consortia (153). Accumulating products included benzoic acid, ethylbenzene, toluene, p-cresol, and benzyl alcohol. The disruption of interspecies hydrogen transfer and its effect o reduced product formation could mimic microbial processes in the subsurface where fermentative conditions are established but methanogenesis is otherwise inhibited.

Transformations of TEX compounds were found in both laboratory microcosms and filed sites at the Sleeping Bear site contaminated with alkylbenzenes from a petroleum spill (83). Benzene was recalcitrant in field and laboratory. Toluene was utilized preferentially to other TEX compounds. A groundwater contaminant plume downgradient to a municipal solid waste landfill was monitored for benzene, ethylbenzene, toluene, and the three xylene isomers. Toluene was utilized relatively rapidly in the methanogenic/sulfate reducing zone of the plume, while benzene and ethylbenzene persisted (150, 151). The xylenes were degraded slowly in the methanogenic/sulfate reducing zone, leading to an increase in the ethylbenzene to xylene ratio with downgradient distance.

Laboratory microcosm studies with methanogenic aquifer material from Traverse City, MI indicated that BTU monoaromatic hydrocarbons declined by an order of magnitude in eight weeks time. Wellwater samples from methanogenic zones of the plume showed degradation products similar to those found in laboratory studies of methanogenic systems subjected to BTEX compounds. Compounds detected in methanogenic aquifer samples included the three cresols, benzoic acid, 2- and 4-methylbenzoic acid, 2,3- and 3,5-dimethylbenzoic acid, phenol, and 2,4 dimethylphenol.

Groundwater monitoring of the dissolved constituents of the methanogenic core of the hydrocarbon plume at Bemidji, Mn, indicate the presence of numerous organic acid degradation products from microbial utilization of benzene and 1C to 4C alkylbenzene parent compounds (85). Phenol, aromatic acids, alicyclic acids. and straight- and branched-chain aliphatic organic acid products were detected. Acetic acid was the predominant aliphatic organic acid; its concentration in the downgradient direction increased and then decreased, potentially corresponding to acidogenesis and subsequent methanogenesis from acetate. The aromatic acid intermediates become a more significant fraction of the total acid pool downgradient, as parent compounds are transformed through these intermediates (40). Many degradation products and non-degrading alkylbenzenes were rapidly utilized when downgradient movement brought the groundwater into contact with molecular oxygen. Additionally, outgassing of methane indicated that methanogenic reactions were occurring where alkylbenzenes were being transformed (127). These studies indicate that intermediates produced during fermentation of alkylbenzenes and alkylbenzoic acids are a significant component of the overall process of hydrocarbon mineralization; the high accumulation of acetate suggests that the aceticlastic methanogenic reaction may limit the overall rate of hydrocarbon degradation in anaerobic environments. At this site, an evolution from manganese reducing to

methanogenic and iron reducing microbial processes has been observed as a primary attenuation mechanism for BTEX compounds (154).

Oxygenates added to gasoline have been shown to be biodegradable under methanogenic conditions. Methyl tertiary butyl ether (MTBE), ethyl tertiary butyl ether (ETBE) and tertiary butyl alcohol (TBA) were degradable under methanogenic conditions, with TBA having faster degradation rates and MTBE the slowest (62). Degradation of both all substrates was inhibited by addition of ethanol, which may have been preferentially utilized by methanogenic processes. MTBE and ETBE degradation were found only in soils with low organic carbon content and at pH 5 to 6.

Summary tables of reported transformations of organic compounds of petroleum origin under denitrifying, sulfate reducing, and methanogenic conditions are included in the full report submitted to Armstrong Laboratory.

Structured Biochemical Modeling of Alkylbenzene Biodegradation

When Air Force jet fuels are released into the subsurface, chemical components within the fuel may undergo microbial transformations. Biotransformation causes a change in chemical structure of the individual chemical components of the fuel and a change in toxicity of the mixture. In active bioremediation processes, molecular oxygen is often supplied to stimulate aerobic biodegradation of fuel components in the subsurface. The rates of biodegradation of jet fuel components are generally higher when oxygen is available as an electron acceptor than under anaerobic conditions. In addition to slower biodegradation rates, intermediate metabolites are often found that are products of partial oxidations of the parent compounds in jet fuels. Intermediate products include numerous carboxylated organic acids that are presumably derived from the parent alkylbenzenes originally present in the jet fuel.

Anaerobic transformations of jet fuel components have perhaps a greater tendency to lead to the formation of intermediate compounds than do aerobic processes. This is partially due to the complex nature microbial transformations under anaerobic conditions. Under anaerobic conditions, complete mineralization of even relatively simple organic compounds may require the combined activities of multiple microbial species. In addition, anaerobic organisms often gain a small amount of free energy from catabolism of specific substrates, and growth rates of individual members or a consortium and of the culture as a whole may be low. Intermediates may accumulate if transformation of a metabolite product is slower than formation of the metabolite from its precursor. Intermediate accumulation could be a transient condition that is ameliorated with time as microbial activity further develops. Conversely, if a groundwater plume of jet fuel components is convected past a given point in the subsurface, metabolites may form and migrate downgradient. In this case, the migrating metabolites would contribute to the toxicity of the plume. If metabolites formed in anaerobic zones were degraded at faster rates under aerobic conditions, then the entry of the plume into a downgradient aerobic zone could contain the migration of metabolite toxicity. The downgradient aerobic zone could be due to intrinsic biological processes or to a more aggressive control technology such as a bioreactive barrier.

Microbial transformations of jet fuel components under anaerobic conditions are not well understood, particularly when nitrate is not available as an alternate electron acceptor. First order biodegradation rate constants are often applied in subsurface fate and transport models; these are often extrapolated from field data or microcosm studies. In this approach, the overall biological activity is lumped into a single rate constant. To more fundamentally understand biodegradation rates, a more detailed description of microbial transformations is necessary. Though many studies have documented anaerobic microbial transformations of alkylbenzenes, most have been accomplished in batch or column laboratory studies, or using field sites data. The elucidation of growth and substrate utilization parameters in continuous culture have not been accomplished. Indigenous microorganisms, though present in the subsurface in initially small numbers and usually attached to the particle surfaces, can grow and increase their mass and activity in response to a convected subsurface flow of jet fuel components. An understanding of the transformation of parent compounds, metabolite formation and utilization, and mineralization to inorganic products requires a detailed description of the interacting microbial activities. Few studies have attempted to impose biochemical structure on alkylbenzene degradation in the subsurface (202,207,210).

A mathematical model was developed to predict substrate and electron acceptor utilization rates, product formation, the change in water quality due to anaerobic microbial activity in the subsurface. The intent of the model is to predict the rate of utilization of parent compounds and metabolites in the presence and absence of nitrate, ferric iron, and/or sulfate as external electron acceptors. Methanogenesis was also included. The initial model considered only benzoate as an immediate metabolite from toluene biodegradation; other metabolites such as benzylsuccinic and benzylfumaric acids were not included in this initial approach. The modeling approach could be extended to include these intermediates as well as numerous metabolites from more complex alkylbenzenes (Section IV).

The model reactions considered are listed in Table 4, as are the microbial populations and the reactions mediated by each. The structure of the model allowed toluene mineralization, for example, to be mediated by one or multiple microbially mediated reactions depending on the availability of external electron acceptors and the presence or absence of an indigenous population of microorganisms capable of mediating a particular reaction. The stoichiometry of each biochemical reaction was specified by considering the conversion of electron equivalents of donor to catabolic end products and to synthesis. The energy made available from the catabolic reaction was coupled to the energy required for synthesis according to the model of McCarty (94). The rate of substrate utilization for catabolism was expressed as the product of the active cell mass mediating each reaction and the maximum electron transport rate for catabolism. Monod kinetics was for electron donor and for the electron acceptor as well if one were used in the catabolic reaction. If an organic compound was the electron donor for a catabolic reaction, the same organic compound was also the electron donor and carbon source for synthesis. If molecular hydrogen was the electron donor in the catabolic reaction, H_2 and carbon dioxide were respectively the electron donor and carbon source for synthesis.

The model can be applied in three different modes. The first mode is as a batch reactor, with no flow of components or organisms into or out of the reactor. This mode corresponds to microcosms commonly used to study bioremediation processes in collected field soils. The second mode includes influent and effluent flow for both

chemical components and microorganisms. This model mode corresponds to a hemostat if the influent microbial biomass concentration is zero, or to bioaugmentation in suspended growth reactors if influent microbial population is greater than zero. The third mode considers flow of components into and out of the reactor, but complete retention of biomass. This mode can be used to examine the case where microorganisms grow attached to soil particles, but metabolize substrates that are being convected by groundwater flow. This model mode most closely corresponds to subsurface bioremediation where contaminants, when convected from a source area, are used by initially small indigenous microbial populations which multiply in response to the substrate and eventually reduce the substrate concentration. Intrinsic bioremediation is an example of this case.

The biochemical model was applied by specifying influent concentrations of chemicals entering a control volume corresponding to a subsurface reaction volume. Initial chemical and microbial population levels were also selected for the control volume. Active microbial populations were selected by specifying initial non-zero concentrations of the individual microbial in the control volume. An initial population level of zero precluded development of the microbial population and the reaction it mediated. the model was employed in a flow mode for components and batch mode for microorganisms. This corresponds to a contaminant plume migrating through a control volume which supports the development of microbial species that are attached to the aquifer solids and not moving with the groundwater flow.

An example simulation was conducted for an inflowing water containing 0.0003 M toluene, 0.001M Fe(III) and 0.0004M sulfate. Initial levels of toluene were zero, and initial Fe(III) and sulfate levels equaled their influent concentrations. The active microbial populations used in the simulation were 3, 5 to 7, 9 to 11, and 13 to 15, each at an initial concentration of 10^{-4} grams volatile suspended solids per liter of pore water. This initial population level corresponds to coverage by a one micron thick layer of microorganisms of 0.01% of the surface area, assuming the soil to be uniform spherical grains of 1 mm diameter. The model thus simulated the time course of substrate and product concentrations, electron acceptors, and microbial population development when Fe^{+3} and sulfate were available as external electron acceptors. Methanogenesis was also a potentially active process because initial, non-zero concentrations were specified for populations mediating reaction 11 (aceticlastic methanogenesis) and reaction 15 (hydrogenotrophic methanogenesis).

Model outputs are shown in Figures 1 through 4. Figure 1 shows the time course of toluene increase in the volume element over 100 days of simulation. Toluene increased to over 20 mg/l in the control volume, while the degradation intermediates benzoate and acetate increases were relatively minor. Figure 2 shows that iron was the preferred electron acceptor as evidenced by its decreasing concentration, while sulfate was not utilized over the 100 day simulation time. This is because of the greater amount of free energy available from microbial iron reduction results in higher growth rates than for the sulfate reducers. The simulation also predicted no methanogenesis during the 100 day simulation, even though the starting population of methanogens was the same as all other organisms. The microbial population predictions in Figure 3 show that iron reducing organisms grew significantly faster than other species. The specific growth rates of the microbial populations shown in Figure 4 show that the rates of growth vary with the availability of electron acceptors. Negative specific growth rates indicate a declining

microbial population due to endogenous decay. Further development of biochemically structured models is needed to enhance the quantitative understanding of microbial processes to support bioremediation application and risk reduction.

Table 4
Structured Biochemical Model of Anaerobic Toluene Degradation

Reaction	Microbial Population	Catabolic Electron Donor	Catabolic Electron Acceptor	Products
1	1	Toluene	NO_3^-	N_2
2	2	Toluene	Fe^{+3}	Fe^{+2}
3	3	Toluene	-	Benzoic Acid, H_2
4	4	Benzoic Acid	NO_3^-	Acetate, N_2
5	5	Benzoic Acid	Fe^{+3}	Acetate, Fe^{+2}
6	6	Benzoic Acid	SO_4^{-2}	Acetate, HS^-
7	7	Benzoic Acid	-	Acetate, H_2
8	8	Acetate	NO_3^-	N_2
9	9	Acetate	Fe^{+3}	Fe^{+2}
10	10	Acetate	SO_4^{-2}	HS^-
11	11	Acetate	-	CH_4
12	12	H_2	NO_3^-	N_2
13	13	H_2	Fe^{+3}	Fe^{+2}
14	14	H_2	SO_4^{-2}	HS^-
15	15	H_2	CO_2	CH_4
16	15	Formate	-	H_2

Chemical Properties of Jet Fuel Components and Metabolites

Petroleum products such as JP-4 consist of a very large number of individual chemical components with different chemical properties. When jet fuels are released into the subsurface, the available terminal electron acceptors and the specific chemical structure can influence the rate of biodegradation and metabolic product formation from individual compounds. Human and environmental risk can be caused by both the chemicals originally present in jet fuels and by metabolic products not originally present when the fuel was released into the environment. Knowledge of the fate and transport of the parent compounds and their metabolic products is required to access the risk present from subsurface spills.

Often, biotransformation introduces oxygen into the chemical structure, rendering the products of transformation more water soluble and with greater potential for migration in the subsurface. If the metabolites have greater potential for migration, they could move away from the regions in which they are produced faster than

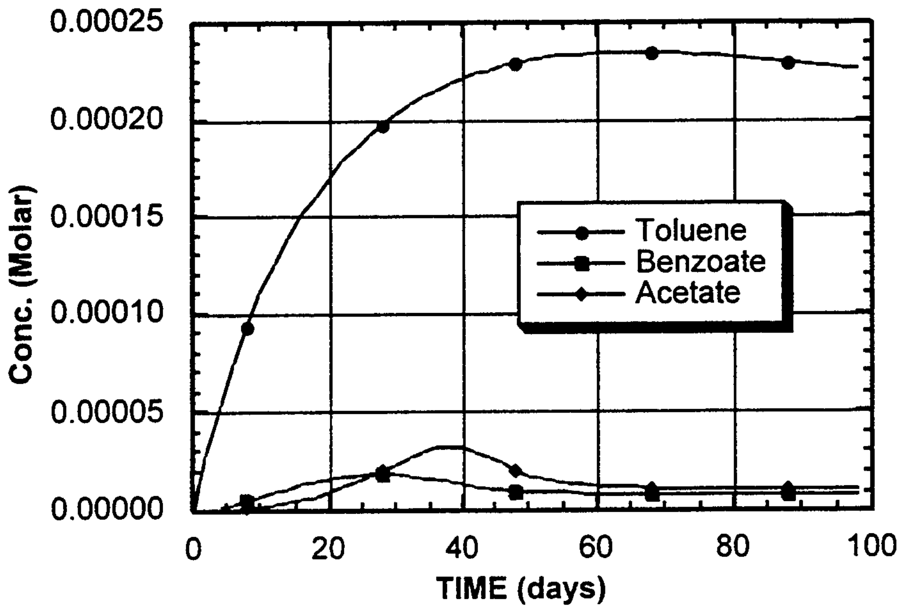


Figure 1 Toluene, Benzoic Acid, and Acetate Concentrations (0.001M Toluene in Influent)

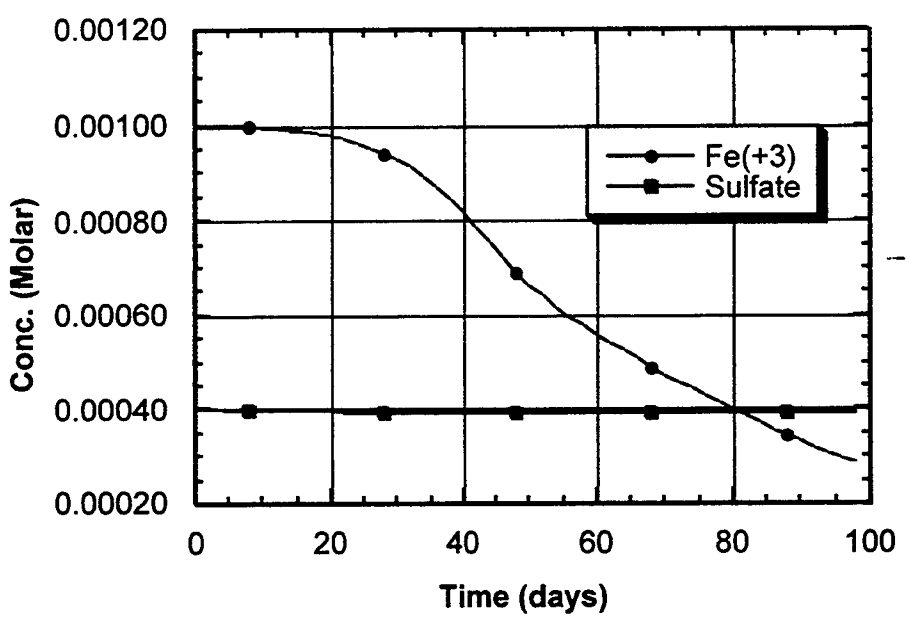


Figure 2 Ferric Iron and Sulfate Concentrations (0.001M Toluene in Influent)

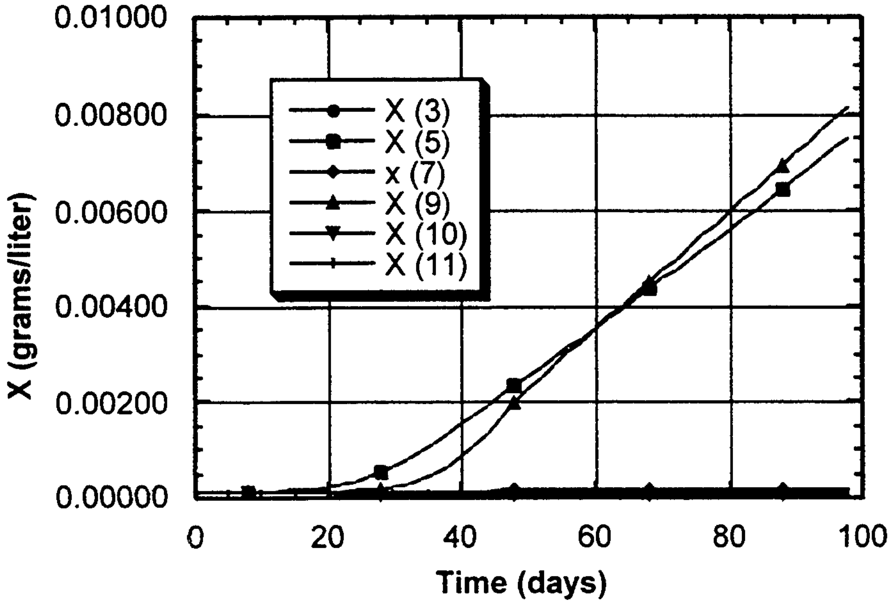


Figure 3 Microbial Population Development
(0.001M Toluene in Influent)

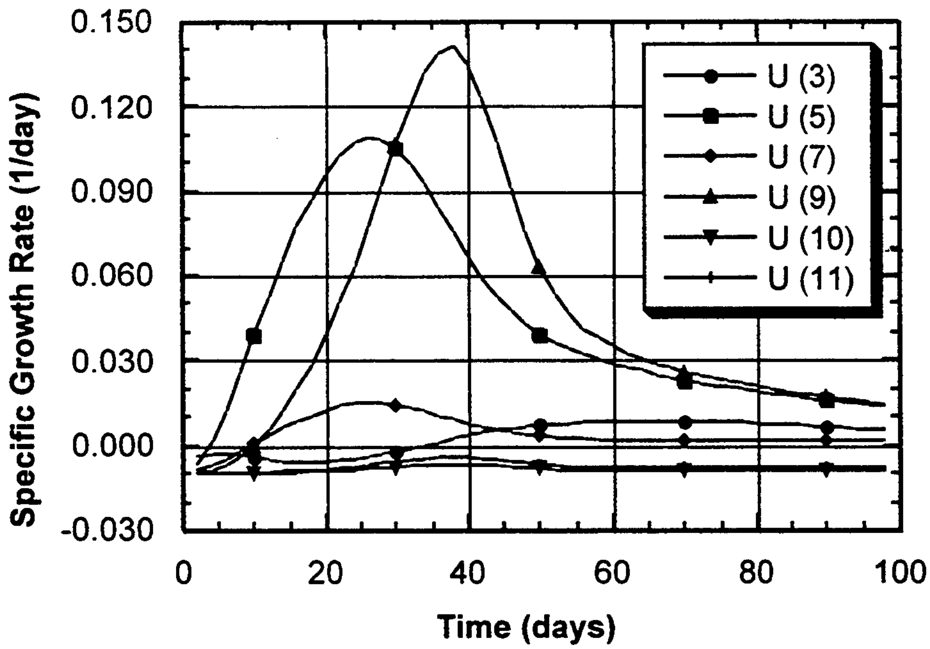


Figure 4 Specific Growth Rates of Microbial Populations

they are degraded by microbial activity. In this case, metabolite toxicity could extend to large spatial regions of the subsurface even if the parent compounds are not present. Examination of the human health and ecological risks from metabolites formed from anaerobic biodegradation of Air Force jet fuels is therefore warranted.

Anaerobic transformations often involve the sequential activities of multiple microbial species to affect complete mineralization of an organic compound, and have perhaps a greater potential for the formation of microbial metabolites than if oxygen is present. Intermediate metabolites may accumulate in the subsurface if they are formed faster than they are utilized. This could occur for example if a higher energy terminal electron accepting process is required for degradation of a metabolite than for its formation. In this case, metabolites formed in anaerobic zones could migrate to downgradient aerobic zones where they could be then consumed. Another situation is where parent compounds are degraded by syntrophic relationships between multiple microbial species. If the subsurface microbial utilization of a metabolite develops slower than utilization of a precursor chemical compound, the metabolite could initially accumulate but then be utilized as the microbial activity necessary for complete mineralization develops. Toluene transformation in methanogenic systems is believed to have a syntrophic nature, and could provide an example of multiple species interactions.

Estimates of chemical properties are needed to assess the potential for migration of metabolites formed during intrinsic and engineered bioremediation processes and to assess the bioavailability and toxicity of intermediates and residuals. Chemical properties of interest include the octanol-water partition coefficient, solubility in water, and the Henry's Law constant. The octanol water partition coefficient is used to predict the binding of neutral hydrophobic organic compounds to organic matter in subsurface sediments. A highly water soluble compound has a greater ability to migrate in the subsurface than a low solubility chemical. The Henry's Law constant is an air water partition coefficient, often equivalent to the ratio of the pure component vapor pressure to the water solubility. A high Henry's Law constant is associated with a increased tendency to volatilize.

To estimate the chemical properties of JP-4 parent compounds and metabolites, Group Contribution Methods (GCMs) contained within the Data Evaluation System for Organic Compounds (DESOC) were used (171). Octanol water partition coefficient, water solubility, and Henry's Law constants were generated for an array of components of Air Force JP-4 fuels and potential microbial metabolites.

The results of GCM application indicate that K_{ow} and K_{aw} decrease and increase, respectively, as the parent compound is transformed to carboxylated metabolites. The introduction of oxygen in carboxylic groups generally increases water solubility, although for some components there is a marked differences in the water solubility predicted by the two GCMs. Higher water solubility could lead to greater toxicity from metabolic intermediate structures as bioremediation progresses because metabolites have greater bioavailability. Such toxicity could decline as the intermediates are then utilized. Toxicity monitoring at several bioremediation sites has shown an initial increase in toxicity during bioremediation followed by a subsequent decrease. These toxicity patterns could be related to changes in the chemical properties of components during bioremediation.

Effects of Bioremediation on Toxicity

Risk assessments at sites contaminated with petroleum hydrocarbons are hampered by the multi-component nature of the contaminants present (275-277). In order to appreciate the effects of bioremediation on toxicity, the biodegradation of specific components of the mixture must be examined and the specific toxicological properties of the remaining components assessed. In addition, the degree to which contaminants are biologically available to be degraded by microorganisms and to exert toxicity are significant questions in groundwater systems. A central concept in toxicology is that the toxicity of a substance is due to the internal dose received by the organism. In soil systems, bioavailability can be reduced by sequestering of contaminants to the high surface areas. Weissenfels et al. found that PAH toxicity was significantly reduced due to adsorption on soil particles; availability to biological degradation was also reduced (248)

Peterson (269) presented methods to calculate the acute toxicity of hydrocarbon mixtures to algal cultures based on the assumption that individual hydrocarbons are equally toxic on the basis of internal concentration within the organism. Results were presented for a series of alkanes, alkylbenzenes, napthalenes, and other chemicals. This viewpoint holds that the differences in measured acute toxicities are due to differences in their equilibrium partitioning between water and the organism. The assumptions of equipotency and additivity could be invoked for neutral hydrophobic narcosis and potentially extended to higher life forms. Peterson accounted for bioavailability by performing equilibrium partitioning calculations to relate predicted dissolved aqueous concentrations toxicity. Quantitative structure activity relationships (QSARs) for non-polar toxicity can be used to generalize, systematize, and extend experimental results to many species. It is not clear if equilibrium partitioning models have been used to calculate actual aqueous concentrations of single hydrocarbons or mixtures in many of the toxicology studies available in the literature (270-274).

A variety of new and established tests have been used to assess the toxicity of various environmental media (260-268). Wang found that the three 'base set' toxicity tests (fathead minnow, macroinvertebrate, and green alga) do not provide adequate characterization of ecotoxicity to higher plants (258). Results were presented comparing millet to fathead minnow toxicity when normalized using structure activity relationships. Short term bacterial tests for the detection of genotoxic agents are evolving rapidly (259). Important progress has been made in understanding the chemical nature of DNA lesions, enzymatic processing of DAN lesions, and the mechanisms of mutagenesis, and test batteries of bacterial tests have been proposed to evaluate the carcinogenicity of chemical compounds (259). One such test, the modified SOS-chromosome procedure, has been applied to test for genotoxicity and cytotoxicity in aquatic sediments without extraction (265). The SOS-chromosome test is a potential substitute for the use of higher organisms such as earthworms and benthic invertebrates (265). Mutatox, a new mutagenic bioassay, has been applied to direct chemical fractionation of organic contaminants in an estuarine sediment (263). The Salmonella/microsome mutagenic assay was used to direct the chemical analysis of genotoxic components in coastal sediments (264).

Risk assessment following bioremediation at sites contaminated with petroleum hydrocarbons requires a knowledge of which components are removed during the bioremediation process and the specific toxicity of each component or class of components. Studies have addressed the first need of identifying the specific chemical composition of hydrocarbon contaminants and changes during bioremediation processes (227,228,230,231,234). It is well known that the rate and extent of biotransformation depend on the chemical structure (219,220,223-226,232,233,235,237-241). Husemann presented a predictive model for estimating the extent of biodegradation of petroleum hydrocarbons in contaminated soils (227). Using a comprehensive petroleum hydrocarbon characterization procedure involving group type separation analyses, boiling point distributions, and mass spectroscopy, initial and final concentrations of specified hydrocarbon classes were determined during seven different bioremediation treatments. It was found that the degree of biodegradation of total petroleum hydrocarbons (TPH) was mainly affected by the chemical structure of the specific components of the mixture, and not by environmental variables. Husemann was able to predict the extent of TPH biodegradation from the average of 86 individual hydrocarbon classes and their respective initial concentrations. Although toxicity testing was not performed, this study provided a systematic analytical/classification method for examining the reductions of specific petroleum hydrocarbon components. Such a method, if coupled with measurements of metabolite formation and toxicological investigations, would provide insight into the effects of bioremediation on the toxicity of mixtures.

Other approaches to the study of biodegradation of petroleum hydrocarbons have used, in addition to TPH reduction, methods such as metabolite formation (222) and computerized mass spectrometry (229). These studies have demonstrated the complex chemical structures that are formed from biodegradation processes; these complex metabolites are not considered in advanced fate and transport models (236).

Several published studies have specifically measured changes in toxicity of media as a result of bioremediation processes. Wang and Bartha performed biodegradation experiments on jet fuel and heating and diesel oils using outdoor lysimeters; they a good correlations between residue decline and toxicity reduction. Toxicity was assesses by Microtox, seed germination, and plant growth bioassays (244,245,251). In another study, bioremediation was applied to soil contaminated by a diesel spill (254). TPH and polyaromatic hydrocarbon persistence were decreased, and residual mutagenicity and acute toxicity, assessed by the Microtox and Ames test, mirrored the decrease chemical components. After substantial initial mutagenicity and toxicity, the contaminated soil approached background levels of uncontaminated soils after 12 weeks of bioremediation.

Carroquino, et al. applied the Ceriodaphnia acute toxicity test to determine the degree of toxicity reduction associated with bioremediation of gasoline contaminated groundwaters under denitrifying conditions, and compared these results with aerobic bioremediation (255). The majority of toxicity from contaminated groundwaters was removed when aeration was used to strip the volatile components, suggesting that the non-volatile organic contaminants contributed only slightly to the toxicity. It was found that bioremediation under nitrate reducing conditions was nearly as effective as aerobic bioremediation in reducing toxicity in the groundwater samples (255).

Reduction in genotoxicity has been reported following fungal bioremediation of a creosote contaminated soil by the Tradescantia-micronucleus test (257). Soil extracts before bioremediation exhibited a strong genotoxic

effect even at a 1% concentration. A decrease in soil genotoxicity was associated with depletion of polyaromatic hydrocarbons following fungal bioremediation. When soil samples were incubated without fungal inoculation, an increase in genotoxicity was observed, and was thought to be due to the generation of water soluble metabolic intermediates by indigenous microflora. Another study assessed toxicity reduction of creosote contaminated soil by fungal based bioremediation using assays of higher plants (250). Both seed germination and root elongation tests showed significant detoxication of soil which correlated well with parent compound depletion.

Mueller et al. examined the rate and extent of biodegradation of pentachlorophenol (PCP) and creosote at a contaminated site, and performed a toxicity assessment of the remediation process. Microtox assays, fish toxicity tests, and teratogenecity tests were used to assess detoxication of the contaminated soil. After two weeks of bioremediation treatment, substantial removals of measured phenolic and lower molecular weight PAH were observed, but only 53% of higher molecular weight PAH and no PCP removal. Despite the removal of the majority of the organic contamination through biotreatment, only slight decreases in toxicity and teratogenicity were observed. The data suggested that toxicity and teratogenicity were associated with compounds that were difficult to treat biologically.

Gersberg et al. used the Ceriodaphnia dubia acute test to examine changes in toxicity in a in a gasoline contaminated aquifer undergoing in situ bioremediation by nitrate addition (252). Substantial reductions in toxicity were evidenced by increased LC50s following bioremediation. However, the results demonstrated that even after bioremediation of an aquifer, with an associated BTEX reduction of 81 to 99%, the toxicity of the groundwaters may not be reduced to precontamination levels.

Pothuluri et al. examined fungal detoxification of fluoranthene *by Cunninghamella elegans* and investigated the mutagenic activity of five metabolites (242). Mutagenic activity of the metabolites was less than fluoranthene , and the mutagenic activity of incubation extracts decreased with time as bioremediation proceeded. *Salmonella typhimurium* strains were used in the mutagenicity assays. These studies are of interest because PAHs, like other chemical carcinogens, exert their carcinogenicity by oxidative metabolism to reactive intermediates (248).

Hund et al. presented an ecotoxicity evaluation strategy and exemplified its use-to assay toxicity reductions from bioremediation of a PAH contaminated site (243*). Pseudomonas putida, Photobacterium phosphoreum*, daphnids, algae, and fish toxicity tests were performed. In addition, soil toxicity was assessed using introduced organisms (plants, earthworms) and natural soil organisms (nematodes, microorganisms). In all test systems, a correspondence between decreasing toxicity and degradation of the easily biodegradable PAHs was found. The test with Daphnia magna indicated the formation of organism specific toxic metabolites. The authors conclude that useful information is gained by biological analyses that complement chemical analyses, and recommend a test battery for extensive assessment of a contaminate site.

Matthews et al. presented a toxicity reduction test system to predict the land treatability of hazardous organic wastes (246). The test system employs reduction of acute toxicity exerted by organics in the water soluble fraction of land applied hazardous wastes as the toxicity measurement criteria. Rosenblatt et al. performed a health risk evaluation for a subsurface site contaminated by a diesel spill by assuming removal of 80% of the fractional

mass of total hydrocarbons and increases in the average molecular weight (256). In his analysis, carcinogenic risks and noncarcinogenic hazard indices were calculated based on the estimated chemical properties of the mixture.

Recommendations

Bioremediation has the potential to reduce the human health and ecological risk from groundwaters contaminated by petroleum hydrocarbons. However, a greater understanding of bioremediation processes is needed before the advantages of biotechnology will be fully realized. Scientific and engineering studies should be conducted to expand knowledge of the effects of bioremediation on soil and groundwater and to increase the confidence in risk assessments at contaminated sites.

There is a paucity of information on the specific components of petroleum hydrocarbons that remain in the soil and groundwater as bioremediation proceeds. Residual petroleum components would be expected to have lower water solubility and volatility than components that are removed during bioremediation, making them less mobile but also less amenable to analytical determination. Field and laboratory studies are needed to critically examine the residuals remaining during and after bioremediation processes have been implemented. Such studies should examine and develop extraction, identification, and quantitation procedures during bioremediation. Aerobic bioremediation processes are a logical first choice for this research effort.

Research is needed on the production of metabolites from bioremediation processes and their effect on toxicity reduction. The chemical structures of microbial products, their fate, transport and toxicological properties, and the susceptibility of metabolites to biodegradation are relatively unexamined. Particularly significant may be the influence of the terminal electron accepting process on the formation rates, concentrations, and chemical structures of metabolites. Laboratory column studies using alternate electron acceptors could provide valuable insight and permit the development of working models of contaminant/metabolite interactions. The use of new analytical techniques such as molecular hydrogen concentration to delineate the terminal electron accepting process and relate it to compound degradation, metabolite formation and toxicity reduction. Toxicity testing protocols should accompany column studies to provide insight into the relationships between measured bioremediation parameters and toxicity reduction.

Acknowledgments The help and assistance of Cathy Vogel of Armstrong Laboratory and Erica Becvar of Applied Research Associates is greatly appreciated. I thank the staff of the Armstrong Laboratory Technical Information Center for their invaluable help in conducting the literature search and in procuring reference material.

References are contained in the original report submitted to Armstrong Laboratory: *Subsurface Bioremediation of Hydrocarbons and Its Effect on Toxicity (1996).*

JOINT CORRECTIONS FOR CORRELATION COEFFICIENTS

Joseph M. Stauffer
Department of Management and Finance

School of Business
Indiana State University
Terre Haute, IN 47809-5402

Final Report for:
Summer Faculty Research Program
Armstrong Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, DC

and

Armstrong Laboratory

August 1996

JOINT CORRECTIONS FOR CORRELATION COEFFICIENTS

Joseph M. Stauffer
Department of Management and Finance
Indiana State University

Abstract

Corrections for range restriction and unreliability are common in psychometric work. Current methods for applying these corrections jointly fail to take into consideration the potentially harmful impact one correction has on the conditions necessary for making the other correction. Using classical test theory, we derive new joint correction formulas that avoid this problem and show how joint correction as currently practiced is sometimes inappropriate.

JOINT CORRECTIONS FOR CORRELATION COEFFICIENTS

Joseph M. Stauffer

Basic formulas for correcting correlation coefficients for unreliability and range restriction have been available for most of this century (Pearson, 1903; Spearman, 1904) and, consequently, are well known and frequently applied. Many applications such as test validation, validity generalization, and psychometric meta-analysis attempt to use both types of correction in conjunction with one another (Hunter & Schmidt, 1990; Mendoza & Mumford, 1987). However, because the correction formulas for range restriction and unreliability were derived separately, it is possible for one correction to alter the conditions necessary to apply the other correction. To avoid this possibility, careful attention must be paid to the sequence with which these corrections are made. Under current practice, the correction sequence is determined using this basic rule-of-thumb: <u>If the reliability coefficient is itself restricted, correct the correlation for unreliability before correcting for range restriction; otherwise, correct the correlation for range restriction and then for unreliability</u>. Using the cases where (1) range restriction is imposed by direct selection on an observed variable and (2) by selection on its latent variable, we derive formulas for joint correction. These formulas show that the current rule-of-thumb is inadequate for determining the correction sequence. The correction sequence is properly determined by the nature of the range restriction, not the nature of the available reliability estimate.

Deriving Joint Correction Formulas

We begin by defining $x = t + e$, where x is an observed variable and t and e are its latent components. The latent variable t is the true score component, and e represents random measurement error. Using uppercase letters to indicate unrestricted values, we define the unrestricted correlation between t and some variable, y, as the parameter of interest:

$$P_{ty} = \frac{\Sigma_{ty}}{\Sigma_t \Sigma_y}, \tag{1}$$

where y can be either a latent variable or an observed variable, Σ_{ty} is the unrestricted covariance between t and y, and Σ_t and Σ_y are the unrestricted standard deviations of t and y. We further assume that $\Sigma_e^2 > 0$ so that $\Sigma_x^2 > \Sigma_t^2$.

The observed correlation, which is restricted and unreliable, is defined as

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}, \tag{2}$$

where σ_{xy} is the restricted covariance between x and y and σ_x and σ_y are the restricted standard deviations of x and y.

To make a proper joint correction of ρ_{xy} for range restriction and unreliability, we must derive formulas that take into account simultaneously the unique assumptions necessary for each form of correction. First, we will briefly summarize the formulas for each correction and the assumptions they make. Next, we will derive new formulas that take these assumptions into account simultaneously--one for the case where direct selection is made on x, and one for the case where selection is made on t.

Correction for Range Restriction

Perhaps the best known and most widely applied form of range restriction correction is the formula corresponding to what Thorndike (1949) labeled Case 2. Case 2 describes a situation where both the restricted and unrestricted standard deviations are known for the variable upon which direct selection occurs. Although found in a variety of algebraically equivalent forms, the correction for Case 2 is given by the formula

$$P_{ab} = \frac{\rho_{ab}}{\sqrt{\frac{\sigma_a^2}{\Sigma_a^2}\left(1 - \rho_{ab}^2\right) + \rho_{ab}^2}}, \tag{3}$$

where $\underline{a}$ is the variable upon which selection is made, $\underline{P}_{ab}$ and $\underline{\rho}_{ab}$ are the unrestricted and restricted correlations between $\underline{a}$ and $\underline{b}$ respectively, and Σ_a^2 and σ_a^2 are the unrestricted and restricted variances of $\underline{a}$. This is the model we will be using to derive our joint correction formulas.

This correction formula is derived from two fundamental assumptions. First, the slope of the regression of $\underline{b}$ on $\underline{a}$ is the same in both the restricted and unrestricted spaces. That is,

$$B_{b|a} = \beta_{b|a}, \tag{4}$$

where $\underline{B}_{b|a}$ is the slope in the unrestricted space and $\underline{\beta}_{b|a}$ is the slope in the restricted space. Second, the variance of $\underline{b}$ given $\underline{a}$ is also assumed to be equal in both the unrestricted and restricted spaces:

$$\Sigma_b^2\left(1 - P_{ab}^2\right) = \sigma_b^2\left(1 - \rho_{ab}^2\right). \tag{5}$$

Note that the variables a and b can be either observed or latent variables. When selection is made on x, the ratio σ_x^2/Σ_x^2 is used in Equation 3 to make the correction. When selection is made on t, the ratio σ_t^2/Σ_t^2 is used.

Correction for Unreliability

Again, suppose the correlation of interest, P_{ty}, is defined as

$$P_{ty} = \frac{\Sigma_{ty}}{\Sigma_t \Sigma_y}. \tag{6}$$

The unrestricted correlation between x and y, P_{xy}, is similarly defined. Because under classical test theory $\Sigma_{ty} = \Sigma_{xy}$, the method for transforming P_{xy} to P_{ty} simply needs to replace the observed standard deviation, Σ_x, with the latent standard deviation, Σ_t. That is precisely what traditional corrections for unreliability attempt to do. When the reliability of x is defined as

$$\Phi_x^2 = \frac{\Sigma_t^2}{\Sigma_x^2}, \tag{7}$$

dividing P_{xy} by Φ_x accomplishes this exchange. That is,

$$P_{ty} = \frac{\Sigma_{xy}}{\Sigma_x \Sigma_y} \frac{\Sigma_x}{\Sigma_t}. \tag{8}$$

Therefore, the correction for unreliability is

$$P_{ty} = \frac{P_{xy}}{\Phi_x}. \tag{9}$$

Joint Correction with Selection on x

We now derive a joint correction formula to obtain P_{ty} from ρ_{xy} when selection is made on x. For this, we will need to define another measure of t, x', where x' = t + e',

with e' representing random measurement error. We start with the assumption that the restricted and unrestricted slopes from the regression of y on x are equal, that is,

$$B_{y|x} = \beta_{y|x}, \tag{10}$$

and the assumption that the unrestricted and restricted variances of y given x are equal:

$$\Sigma_y^2\left(1 - P_{xy}^2\right) = \sigma_y^2\left(1 - \rho_{xy}^2\right). \tag{11}$$

Equation 10 can be written as

$$\frac{\Sigma_{xy}}{\Sigma_x^2} = \frac{\sigma_{xy}}{\sigma_x^2}. \tag{12}$$

Therefore,

$$\Sigma_{ty} = \sigma_{xy}\frac{\Sigma_x^2}{\sigma_x^2}. \tag{13}$$

To calculate Σ_t, we make an assumption similar to Equation 4 such that:

$$B_{t|x} = \beta_{t|x}, \tag{14}$$

and, therefore,

$$\frac{\Sigma_{xt}}{\Sigma_x^2} = \frac{\sigma_{xt}}{\sigma_x^2}. \tag{15}$$

Because $\Sigma_{xt} = \Sigma_t^2$ and $\sigma_{xt} = \sigma_{xx'}$, the restricted covariance of x and x',

$$\Sigma_t = \sqrt{\rho_{xx'}\frac{\sigma_{x'}}{\sigma_x}\Sigma_x^2}, \tag{16}$$

where $\rho_{xx'}$ is the restricted correlation between x and x'. Note that because direct selection on x forces a negative correlation between t and e, $\sigma_{xx'}$ does not equal σ_t^2. Rather, $\sigma_{xx'} = \sigma_t^2 + \sigma_{te}$, where $\sigma_{te} < 0$ (Mendoza & Mumford, 1987). Therefore, ϕ_x^2 will not make the proper correction. It will tend to undercorrect, since $\phi_x^2 > \rho_{xx'}$.

Now we obtain the third component, Σ_y. As a result of Equation 10, we can show that

$$P_{xy} = \rho_{xy} \frac{\sigma_y}{\sigma_x} \frac{\Sigma_x}{\Sigma_y}. \tag{17}$$

Substituting this result into Equation 11, we find that

$$\Sigma_y = \sqrt{\sigma_y^2 - \sigma_y^2 \rho_{xy}^2 + \sigma_y^2 \rho_{xy}^2 \frac{\Sigma_x^2}{\sigma_x^2}}. \tag{18}$$

Rewriting Equation 13 in terms of ρ_{xy}, substituting that result along with Equations 16 and 18 into Equation 1, and rearranging, we have the joint correction formula

$$P_{ty} = \frac{\rho_{xy}}{\sqrt{\rho_{xx'} \frac{\sigma_{x'}}{\sigma_x} \sqrt{\frac{\sigma_x^2}{\Sigma_x^2}\left(1 - \rho_{xy}^2\right) + \rho_{xy}^2}}}. \tag{19}$$

If y represents an unreliable, observed measure, we would define it as $y = u + f$, where u is the true-score component and f is the error component. In terms of the unrestricted reliability of y,

$$\Sigma_u = \Phi_y \sigma_y \sqrt{1 - \rho_{xy}^2 \left(1 - \frac{\Sigma_x^2}{\sigma_x^2}\right)}. \tag{20}$$

Assuming that the measurement error variance associated with y is unaffected by selection on x, that is, $\Sigma_f^2 = \sigma_f^2$, we define $\Sigma_f^2 = \sigma_y^2(1 - \phi_y^2)$, where $\phi_y^2 = \sigma_u^2/\sigma_y^2$. This allows us to express Σ_u in terms of the restricted reliability:

$$\Sigma_u = \sigma_y \sqrt{\phi_y^2 - \rho_{xy}^2 \left(1 - \frac{\Sigma_x^2}{\sigma_x^2}\right)}. \tag{21}$$

With the classical assumption, $\Sigma_{tu} = \Sigma_{xy}$, we now have an expression for P_{tu} based entirely on observable values. Substituting Equation 20 for Equation 18, we obtain

$$P_{tu} = \frac{\rho_{xy}}{\sqrt{\rho_{xx'}\dfrac{\sigma_{x'}}{\sigma_x}\Phi_y\sqrt{\dfrac{\sigma_x^2}{\Sigma_x^2}\left(1-\rho_{xy}^2\right)+\rho_{xy}^2}}}. \tag{22}$$

By substituting Equation 21 for Equation 18, we have

$$P_{tu} = \frac{\rho_{xy}/\phi_y}{\sqrt{\rho_{xx'}\dfrac{\sigma_{x'}}{\sigma_x}\sqrt{\dfrac{\sigma_x^2}{\Sigma_x^2}\left(1-\dfrac{\rho_{xy}^2}{\phi_y^2}\right)+\dfrac{\rho_{xy}^2}{\phi_y^2}}}}. \tag{23}$$

Joint Correction with Selection on t

When selection is made on the basis of t, Equations 4 and 5 translate to

$$B_{y|t} = \beta_{y|t} \tag{24}$$

and

$$\Sigma_y^2\left(1-P_{ty}^2\right)=\sigma_y^2\left(1-\rho_{ty}^2\right). \tag{25}$$

We may rewrite Equation 24 as

$$\frac{\Sigma_{ty}}{\Sigma_t^2}=\frac{\sigma_{ty}}{\sigma_t^2}. \tag{26}$$

Isolating Σ_{ty} and stating in terms of x, we find that

$$\Sigma_{ty} = \frac{\rho_{xy}}{\phi_x}\sigma_x\phi_x\sigma_y\frac{\Sigma_x^2}{\sigma_x^2}\frac{\Phi_x^2}{\phi_x^2}. \tag{27}$$

From Equation 25 we can show that

$$\Sigma_y^2 = \sigma_y^2 - \sigma_y^2 \rho_{ty}^2 + \Sigma_y^2 B_{y|t}^2 \frac{\Sigma_t^2}{\Sigma_y^2}. \tag{28}$$

Due to Equation 24,

$$\Sigma_y^2 = \sigma_y^2 - \sigma_y^2 \rho_{ty}^2 + \Sigma_y^2 \beta_{y|t}^2 \frac{\Sigma_t^2}{\Sigma_y^2}. \tag{29}$$

Converting $\beta_{y|t}$ back to a correlation coefficient yields

$$\Sigma_y^2 = \sigma_y^2 - \sigma_y^2 \rho_{ty}^2 + \rho_{ty}^2 \frac{\sigma_y^2}{\sigma_t^2} \Sigma_t^2. \tag{30}$$

Expressing the equation in terms of x and collecting gives us

$$\Sigma_y^2 = \sigma_y^2 \left[1 - \frac{\rho_{xy}^2}{\phi_x^2} \left(1 - \frac{\Sigma_x^2}{\sigma_x^2} \frac{\Phi_x^2}{\phi_x^2} \right) \right]. \tag{31}$$

The expression for Σ_t would be

$$\Sigma_t = \Sigma_x \Phi_x. \tag{32}$$

Substituting Equations 25, 31, and 32 into Equation 1 and rearranging, we arrive at the joint correction formula for the case where selection is made on t:

$$P_{ty} = \frac{\rho_{xy} / \phi_x}{\sqrt{\frac{\phi_x^2}{\Phi_x^2} \frac{\sigma_x^2}{\Sigma_x^2} \left(1 - \frac{\rho_{xy}^2}{\phi_x^2} \right) + \frac{\rho_{xy}^2}{\phi_x^2}}}. \tag{33}$$

If $y = u + f$, in terms of Φ_y,

$$P_{tu} = \frac{\rho_{xy} / \phi_x}{\Phi_y \sqrt{\frac{\phi_x^2}{\Phi_x^2} \frac{\sigma_x^2}{\Sigma_x^2} \left(1 - \frac{\rho_{xy}^2}{\phi_x^2} \right) + \frac{\rho_{xy}^2}{\phi_x^2}}}. \tag{34}$$

In terms of ϕ_y,

$$P_{tu} = \frac{\rho_{xy} / \phi_x \phi_y}{\sqrt{\frac{\phi_x^2}{\Phi_x^2} \frac{\sigma_x^2}{\Sigma_x^2} \left(1 - \frac{\rho_{xy}^2}{\phi_x^2 \phi_y^2}\right) + \frac{\rho_{xy}^2}{\phi_x^2 \phi_y^2}}}. \qquad (35)$$

Reconsidering Current Practice

The current approach to making joint corrections tells us to correct correlations for unreliability first if our reliability is restricted, and correct for range restriction first, if our reliabilities are unrestricted. This rule-of-thumb corresponds to the following two equations:

$$P_{ty} = \frac{\rho_{xy} / \phi_x}{\sqrt{\frac{\sigma_x^2}{\Sigma_x^2} \left(1 - \frac{\rho_{xy}^2}{\phi_x^2}\right) + \frac{\rho_{xy}^2}{\phi_x^2}}} \qquad (36)$$

and

$$P_{ty} = \frac{\rho_{xy}}{\Phi_x \sqrt{\frac{\sigma_x^2}{\Sigma_x^2} \left(1 - \rho_{xy}^2\right) + \rho_{xy}^2}}. \qquad (37)$$

Equation 36 corresponds to the situation where the reliability of x is restricted (cf., Bobko, 1983; Equation 1, p. 585). Equation 37 corresponds to the case where the reliability of x is unrestricted (cf., Raju, Burke, Normand, & Langlois, 1991; Equation 1, p. 423).

The problem with this approach is that the decisive factor by which we determine the correction sequence is the nature of the reliability of x. Although intuitive, it really

has no mathematical basis. Compare Equations 36 and 37 with our joint correction formulas:

$$P_{ty} = \frac{\rho_{xy}}{\sqrt{\rho_{xx'} \dfrac{\sigma_{x'}}{\sigma_x} \sqrt{\dfrac{\sigma_x^2}{\Sigma_x^2}\left(1 - \rho_{xy}^2\right) + \rho_{xy}^2}}} \tag{38}$$

(Equation 19 above) and

$$P_{ty} = \frac{\rho_{xy} / \phi_x}{\sqrt{\dfrac{\phi_x^2}{\Phi_x^2} \dfrac{\sigma_x^2}{\Sigma_x^2}\left(1 - \dfrac{\rho_{xy}^2}{\phi_x^2}\right) + \dfrac{\rho_{xy}^2}{\phi_x^2}}} \tag{39}$$

(Equation 33 above). These new formulas reveal that it is actually the nature of the range restriction, not the reliability of x, that determines the correction sequence. Equation 38, which, incidentally, is equivalent to Equation 37, shows that when selection is made on the basis of x, the correlation coefficient must be corrected first for range restriction and then for unreliability using the unrestricted reliability, Φ_x^2, where

$$\Phi_x^2 = \rho_{xx'} \frac{\sigma_{x'}}{\sigma_x} \tag{40}$$

when x and x' are parallel, tau equivalent, or essentially tau-equivalent measures of t. Equation 39 shows that when selection is made on the basis of t, the correlation must first be corrected for unreliability using the restricted reliability, ϕ_x^2, and the standard deviation ratio, σ_x/Σ_x, is corrected for unreliability using the reliability ratio, ϕ_x/Φ_x, before the correction for range restriction is made.

The reason for this sequence is quite simple. The correction for range restriction needs to be made on the joint bivariate distribution that includes the variable upon which

selection was made. When selection is made on $\underline{x}$, the range restriction correction requires the correlation $\rho_{\underline{xy}}$ and the ratio, $\underline{\sigma}_{\underline{x}}/\underline{\Sigma}_{\underline{x}}$. Correcting first for unreliability in $\underline{x}$ denies the range restriction correction those requisite parameters. Conversely, when selection is made on $\underline{t}$, the range restriction correction requires the correlation, $\rho_{\underline{ty}}$, and the standard deviation ratio, $\underline{\sigma}_{\underline{t}}/\underline{\Sigma}_{\underline{t}}$. Therefore, we need to correct both the observed correlation, $\rho_{\underline{xy}}$, and the ratio, $\underline{\sigma}_{\underline{x}}^2/\underline{\Sigma}_{\underline{x}}^2$, for unreliability in $\underline{x}$ before applying the range restriction correction. Because the error and true-score components of an unreliable, observed measure, $\underline{y}$, remain uncorrelated under selection on either $\underline{x}$ or $\underline{t}$, $\rho_{\underline{xy}}$ can be corrected for unreliability in $\underline{y}$ either before or after correcting for range restriction.

Consequently, Equation 37, representing the unrestricted reliability condition under the current rule-of-thumb, is appropriate only when the reliability of $\underline{x}$ is unrestricted <u>and</u> selection is made the basis of $\underline{x}$. Equation 36, representing the restricted reliability condition under the current rule-of-thumb, is obviously inappropriate in any situation.

Transforming Reliabilities

Since corrections for unreliability in $\underline{x}$ should no longer be made on the basis of the restriction status of the available reliability estimate (i.e., unrestricted or restricted), we need to be able to transform $\underline{\phi}_{\underline{x}}^2$ to $\underline{\Phi}_{\underline{x}}^2$ and vice versa as the situation warrants. Under the new rules, if selection is made on $\underline{x}$, we correct for unreliability in $\underline{x}$ last. That requires $\underline{\Phi}_{\underline{x}}^2$. If $\underline{\Phi}_{\underline{x}}^2$ is unavailable, we can obtain an estimate of $\rho_{\underline{xx'}}$ and apply Equation

40. It is interesting to note that with the assumption that $\underline{x}$ and $\underline{x}'$ are at least essentially tau equivalent, Equation 40 requires no information at all about the unrestricted space. The transformation is made completely on the basis of restricted parameters or their estimates.

If selection is made on the basis of $\underline{t}$, we need both $\underline{\Phi}_{\underline{x}}^2$ and $\underline{\phi}_{\underline{x}}^2$. If we have an estimate of $\underline{\Phi}_{\underline{x}}^2$, because restriction on $\underline{t}$ does not affect the zero correlation between $\underline{t}$ and $\underline{e}$, we can apply a well-known range restriction transformation to our estimate of $\underline{\Phi}_{\underline{x}}^2$ to obtain $\underline{\phi}_{\underline{x}}^2$:

$$\phi_x^2 = 1 - \left[\frac{\Sigma_x^2}{\sigma_x^2}\left(1 - \Phi_x^2\right)\right] \tag{41}$$

(see, e.g., Lord & Novick, 1968; Equation 6.2.1, p. 130). If instead we have an estimate of $\underline{\phi}_{\underline{x}}^2$, we use the inverse function

$$\Phi_x^2 = 1 - \left[\frac{\sigma_x^2}{\Sigma_x^2}\left(1 - \phi_x^2\right)\right]. \tag{42}$$

Equations 41 and 42 can also be applied to transform the reliability of $\underline{y}$ from one form to another:

$$\phi_y^2 = 1 - \left[\frac{\Sigma_y^2}{\sigma_y^2}\left(1 - \Phi_y^2\right)\right] \tag{43}$$

and

$$\Phi_y^2 = 1 - \left[\frac{\sigma_y^2}{\Sigma_y^2}\left(1 - \phi_y^2\right)\right]. \tag{44}$$

Since we have been assuming throughout that, under Thorndike's Case 2, $\underline{\Sigma}_{\underline{y}}$ and $\underline{\sigma}_{\underline{y}}$ are unknown, we point out that, from Equation 31,

$$\frac{\Sigma_y^2}{\sigma_y^2} = 1 - \frac{\rho_{xy}^2}{\phi_x^2}\left(1 - \frac{\Sigma_x^2}{\sigma_x^2}\frac{\Phi_x^2}{\phi_x^2}\right). \tag{45}$$

Therefore, Equations 43 and 44 are equivalent to

$$\phi_y^2 = 1 - \left[\left(1 - \frac{\rho_{xy}^2}{\phi_x^2}\left(1 - \frac{\Sigma_x^2}{\sigma_x^2}\frac{\Phi_x^2}{\phi_x^2}\right)\right)\left(1 - \Phi_y^2\right)\right] \tag{46}$$

and

$$\Phi_y^2 = 1 - \left[\frac{1 - \phi_y^2}{1 - \frac{\rho_{xy}^2}{\phi_x^2}\left(1 - \frac{\Sigma_x^2}{\sigma_x^2}\frac{\Phi_x^2}{\phi_x^2}\right)}\right]. \tag{47}$$

Summary

The current practice of determining the order in which to apply corrections for unreliability and range restriction according to the nature of available reliability estimates is inadequate. When correcting for unreliability in the variable y, whether latent or observed, the current rule-of-thumb is appropriate. However, because a correction for unreliability in x could adversely affect the conditions necessary for applying the range restriction correction, we must look to the nature of the range restriction to determine the correction sequence. When selection is made on the basis of an unreliable, observed measure, x, we must correct the correlation first for range restriction. When selection is made on the basis of the true-score component of x, t, we must correct the correlation and both the unrestricted and restricted standard deviations of x first for unreliability in x. This means that we must be able to obtain an expression for the restricted reliability of x

when our available estimate of the reliability is unrestricted. Likewise, we need an expression for the unrestricted reliability of $\underline{x}$ when our available estimate is restricted. We presented formulas which accomplish these transformations.

References

Bobko, P. (1983). An analysis of correlations corrected for attenuation and range restriction. Journal of Applied Psychology, 68, 584-589.

Hunter, J. E., & Schmidt, F. L. (1990). Methods of meta-analysis: Correcting error and bias in research findings. Newbury Park, CA: Sage.

Lord, F. M., & Novick, M. R. (1968). Statistical theories of mental test scores. Menlo Park, CA: Addison-Wesley.

Mendoza, J. L., & Mumford, M. (1987). Corrections for attenuation and range restriction on the predictor. Journal of Educational Statistics, 12, 282-293.

Pearson, K. (1903). Mathematical contributions to the theory of evolution. XI. On the influence of natural selection on the variability and correlation of organs. Royal Society Philosophical Transactions, 200 (Series A), 1-66.

Raju, N. S., Burke, M. J., Normand, J., & Langlois, G. M. (1991). A new meta-analytic approach. Journal of Applied Psychology, 76, 432-446.

Spearman, C. (1904). The proof and measurement of association between two things. American Journal of Psychology, 15, 72-101.

Thorndike, R. L. (1949). Personnel selection: Test and measurement techniques. New York: Wiley.

Studies to Identify Characteristic Changes in the Urine Following Ingestion of Poppy seed

William B. Stavinoha, Ph.D.
Professor of Pharmacology
Department of Pharmacology

7703, Floyd Curl Drive
University of Texas Health Science Center
San Antonio, Tx 78284

Final Report for:
Summer Faculty Research Program
Armstong Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, DC

and

Armstrong Laboratory

September 1996

Studies to Identify Characteristic Changes in the Urine Following Ingestion of Poppy seed

William B. Stavinoha, Ph.D.
Professor of Pharmacology
Department of Pharmacology
University of Texas Health Science Center
San Antonio

Abstract

Characteristics of poppy seed that could be used to differentiate the urine resulting from intake of poppy seed from opiate abuse were studied. Meconic acid which is present at 2-13% in opium was not found in the black or white poppy seed studied. Opiate alkaloids other than morphine and codeine were not identified in the urine following poppy seed ingestion. Chronic multiple ingestion of low doses of poppy seed was found to increase the morphine and codeine content of the urine 2-3 fold over single ingestion. The morphine/codeine ratio in chronic and acute ingestion of poppy seed had a range of 4-13 with relatively low variation within individuals. The morphine/codeine ratio appears to provide a tool which may be useful in differentiation of poppy seed consumption from opiate abuse.

Studies to Identify Characteristic Changes in the Urine Following Ingestion of Poppy seed

William B. Stavinoha, Ph D.

Introduction

Analysis of urine is a major tool in the control of consumption of illegal drugs. The use of the combination of the gas chromatograph and mass spectrometer for analysis results in a robust and dependable positive identification and quantitation of the drug and its metabolites in the urine. Use of this powerful analytical system essentially removes doubt on the presence or absence of the drug. Presence of the drug usually indicates illegal use, but there is an exception. Most drugs of abuse do not occur in the normal diet so the identification of that drug in the urine at a predefined level is considered a powerful indicator of illicit use of the compound, however if the drug of abuse is present in the normal diet, it is difficult to establish guidelines for determining whether the source of drug in the urine is the result of drug abuse or normal dietary intake. This is the problem that occurs following ingestion of pastries and other foods containing seeds of Papaver somniferum (poppy seed). Although most poppy seeds do not contain high levels of narcotic alkaloids some poppy seeds can contain as high as 963μg/g of morphine and 79μg/g of codeine. Ingestion of poppy seeds with this high alkaloidal content can result in urinary levels of 17900ng/ml of morphine and 400ng/ml of codeine (Fritschi and Prescott, 1985). When this problem was recognized the level of morphine in the urine that was considered indicative of opiate abuse was raised from 300ng/ml to 4000ng/ml. This high predefined level results in very few positive findings of opiate abuse. A method is needed to differentiate between innocent poppy seed ingestion as a food and abuse of morphine, codeine or heroin.

Discussion of the problem

Research on this problem has focused on three areas of differentiation:

(1) identification of monoacetylmorphine, a metabolic product of heroin, in the urine as an indicator of heroin ingestion(Mule and Casella,1988). A major disadvantage is that monoacetylmorphine is quite unstable and readily hydrolyzed. The detection time range is only 2-8 hours at the most sensitive cutoff limit and readily breaks down before analysis can begin (Cone et al., 1991).

(2) identification of characteristic compounds in poppy seed that could serve as markers of ingestion when they appear in the urine. Poppy seed contains the narcotic alkaloids of Papaver somniferum (Duke, 1985), in low concentration. The drugs of abuse, morphine, heroin and codeine are usually free of these alkaloids unless in used in the very crude form (Yong and Lik,

1977).Fritschi and Prescott (1985) attempted to find the alkaloid narcotoline in the urine following poppy seed ingestion using RIA, enzyme multiple immunoassay techniques (EMIT-ST) and GC, but were unsuccessful. Thebaine metabolites have been found in the urine of monkeys following 8mg/kg sc thebaine (Yamazoe et al., 1981) ElSohly et al (1985) were unable to identify thebaine in the urine following ingestion of 200g of poppy seed cake, but the amount of poppy seed ingested was not stated. The poppy seed used to prepare the cake was analyzed and contained morphine 24ug/g, codeine 0.36 μg/g and thebaine 0.46μg/g.

(3) attempt to identify the source of the alkaloid ingested by comparing the relative amounts of morphine and codeine excreted in the urine as suggested by Yong and Lik (1977) and ElSohly et al.(1990). The conditions to rule out poppy seed ingestion as formulated by ElSohly et al (1990) are: (a) codeine levels exceeding 300ng/ml (b) morphine-codeine ratio of less than 2 (c) 1000ng/ml morphine with no codeine detected (d) morphine levels in excess of 5000ng/ml. (e) presence of 6-monoacetylmorphine.

This research project utilized four approaches. (a) identify a characteristic compound in poppy seed that would be excreted in the urine after poppy seed ingestion, (b) through use of opium alkaloid standards identify these alkaloids in the urine using MSMS following ingestion of poppy seed, (c) to study the increase in levels of morphine and codeine in the urine following multiple ingestions as compared to a single acute ingestion of poppy seed, and (d) to study the ratio of morphine to codeine in the urine following acute and chronic ingestion of poppy seed or poppy seed containing pastries.

(a) Meconic acid occurs in opium to the extent of 2-10% (Annett and Bose, 1922) and 7-13 % (Miyamoto and Brochmann-Hanssen, 1962). Fairbairn and Steele (1981) have reported that meconic acid is restricted to the genus Papaver and some closely related genera of 48 Papaverceous species examined. The deep red color produced by meconic acid with ferric chloride solution is a commonly used test for opium (Lim and Kwok, 1981). Since if present in the body it would be excreted in the urine, meconic acid was chosen as a candidate characteristic compound to measure in poppy seed.

(b) Using standard specimens of the minor alkaloids such as sinomenine, laudanosine, oripavine, narceine and papapervine search, for their presence in the urine following ingestion of poppy seed was carried out on the MSMS instrument.

(c) The effects of acute ingestion of poppy seed pastries on the urinary excretion of morphine and codeine has been extensively reported(Fritschi and Prescott Jr,1985; Hayes et al.,1987; Pettitt et

al.,1987; Struempler,1987; ElSohly et al., 1988; ElSohly and Jones,1989; ElSohly and ElSohly, 1990; Selavka, 1991; Carpenter, 1994; Huestis and Cone, 1995). This data provides information on the maximum excretion of these two alkaloids following a one time ingestion. Realistically, poppy seed ingestion seldom occurs with the eating of one large portion. More likely ingestion will occur with the eating of small amounts of poppy seed containing foods over a period of days following a bakery purchase or baking pastries at home.. If the multiple ingestion rate of the alkaloids exceeds the excretion rate, the concentration of alkaloids will increase in the body and result in an increase in the concentration of the alkaloids in the urine over that from a single ingestion. There have been no reports in the literature on the effects of multiple ingestion of poppy seed extended over several days. This is an important oversight for many ethnic cuisines use poppy seed in a variety of ways. It is important to ascertain the effect on urinary excretion of morphine and codeine of multiple ingestion of poppy seed over several days.

(d) The ratio of morphine to codeine in the urine appears to be a useful but inadequately tested method to identify poppy seed ingestion. Calculation of the ratios of morphine to codeine using the data obtained from papers reporting urinary concentrations of morphine and codeine following poppy seed ingestion indicate that the ratio of morphine to codeine is between 2 and 60 (ElSohly and Jones, 1989). None of the papers were testing the ratio hypothesis and none were done on chronic ingestion. It is necessary for validation to study the hypothesis more completely and to evaluate in many samples including excretion following chronic ingestion of poppy seed.

Methods

Meconic acid standard was obtained from the United Nations International Drug Control Program. Both black and white poppy seed was extracted using several methods to isolate meconic acid for testing. (A). Ten grams of poppy seed in 50 ml of pH 4 citrate buffer was homogenized using a Polytron homogenizer (Brinkman Inst. Westbury, N.Y.) The solution was kept overnight and then filtered. The filtrate was concentrated and redissolved in water. Several drops of dilute HCl were added and several; drops of 5% ferric chloride in 0.1 N HCl were added. Ferric chloride is a classic test for meconic acid producing a red color with solutions containing meconic acid (Lim and Kwok, 1981). (B) Isolation of meconic acid using Dowex resin on the hot water extract of ground poppy seed was also used following the method of Miyamoto and Brochmann-Hanssen (1962).

Standards of alkaloids contained in opium and poppy seed were run on the MS/MS and search for characteristic fragments in urine following poppy seed ingestion were carried out.

Subjects: Three male laboratory members served as the subjects. The age range was 40-68 years. For the acute study raw poppy seeds were ingested at 7:00AM before breakfast. Subject A ingested 47g of poppy seed and subject B ingested 100g of poppy seed In the chronic study poppy seed containing pastries were ingested three times a day for 3 days and total urine output collected. Subjects A and B ingested 13.5 g of poppy seed per dose and subject C ingested 90 g per dose in the three dose study and 15.0 g per dose in the nine dose study.

Both morphine and codeine are excreted in the urine as the glucuronide (Glare and Walsh, 1991; Vree and VanWissen, 1992; Milne et al., 1996); therefore, the urine samples were hydrolyzed before analysis. Morphine/codeine extraction procedure: To 5ml of urine 1ml of 50% HCl was added followed after cooling by 2 ml of 2.0M Tris buffer and 700μl 10N KOH in KHCO3. The mixture was then mixed on a vortex mixer. The pH should be between 8.0 and 9.0. It is then centrifuged at 1500rpm for 2 minutes. The extraction column was cleaned using 2ml methanol followed by 2ml deionized water. The specimen was added to the column and the vacuum adjusted for a minimum 3 minute residence time of the sample. The column was then washed with 2ml deionized water followed by 1ml 100nM acetate buffer (pH 4.0) and then 2 ml methanol. To elute the alkaloids 3 ml methylene chloride:isopropyl:ammonium hydroxide (80:20:2) made fresh daily was used. The eluate was dried at 40 degrees C. The eluate was derivatized with 100μl pyridine, and 100μl acetic anhydride, and incubated for 15 minutes at 70 degrees C. The solution was evaporated at 70 degrees C under nitrogen. The sample was reconstituted with ethyl acetate and analyzed using the GC/MS with deuterated standards for morphine and codeine.

Results

The tests for meconic acid in poppy seed were negative. The meconic acid standard produced a red color at very low concentrations while the poppy seed extracts showed no red coloration.

MS/MS was used to search for fragments in urine following poppy seed ingestion that would match the alkaloid standards. None was found to indicate that any of the minor alkaloids of opium occur in the urine at identifiable levels following poppy seed ingestion at the level of ingestion studied and the instrumentation used.

Multiple ingestion of poppy seed containing pastries results in a large increase in morphine and codeine in the body over single ingestion of the same dose. This results in an increase of the concentration of morphine and codeine excreted in the urine. The figures illustrate the increase in

the excretion In Fig.1 the increase in concentration of morphine and codeine on the third of three times a day ingestion over the highest excretion of the first day of multiple ingestion is for morphine 3.5 fold increase and codeine a 2,7 fold increase. In fig 4 the increase for both morphine and codeine is 1.6 fold and in Fig 7 for morphine 2 and codeine 1.6. Fig 2 which plots total excretion illustrates that codeine excretion is completed 42 hours before morphine. The urine for Fig. 5 and Fig. 8 was not collected long enough to show this excretion difference. The acute ingestion of poppy seed illustrates a strong second excretion peak for both morphine and codeine possibly reflecting enterohepatic recirculation (Dahlstrom and Paalzow, 1978). The concentrations of morphine and codeine are very low and at the edge of sensitivity of the GC/MS analytical method, but are measurable and sequential analyses match closely.

Following ingestion of raw poppy seed before eating at 7AM the ratio of morphine/codeine was 8±2.4 for subject A (Fig 13) and 12±1.3 for subject B (Fig 15). With three doses of cooked poppy seed the ratio was 15±2.7 for subject C. Chronic ingestion of poppy seed containing pastries the ratio for subject A was 4±1 subject B,9±2 and subject C, 13±2.9

Conclusions

1. Meconic acid was not found in either the black or white poppy seed using the ferric chloride test.

2. When urine was extracted by the currently used methods, morphine and codeine were present but the other opium alkaloids other than morphine+codeine were not present in sufficient quantities in urine following poppy seed ingestion to be detected by MS/MS.

3. Ingestion of poppy seed containing pastries three times a day for three days results in an 1.6 to 3.5 fold increase in morphine in the urine and a 1.6 to 2.7 fold increase in codeine in the urine. In the one subject where urine was collected for several days following the ingestion of poppy seed morphine was excreted in the urine for 42 hours after the urine was free of codeine.

4. In two subjects acutely ingesting in one dose the same source raw poppy seed the morphine/codeine ratio was 8±2.4 and 12±1.3. Chronic ingestion three times a day for three days of this same source poppy seed, cooked in a pastry resulted in a ratio of 4±1 and 9±2. Chronic ingestion of a different source poppy seed provided a ratio of 13±2.9 and 15±2.7. This indicates less fluctuation in the ratio than the literature which reports a ratio SD of about 50%.The ratio of morphine/codeine appears to be a useful criteria for identification of both acute and chronic ingestion of poppy seed.

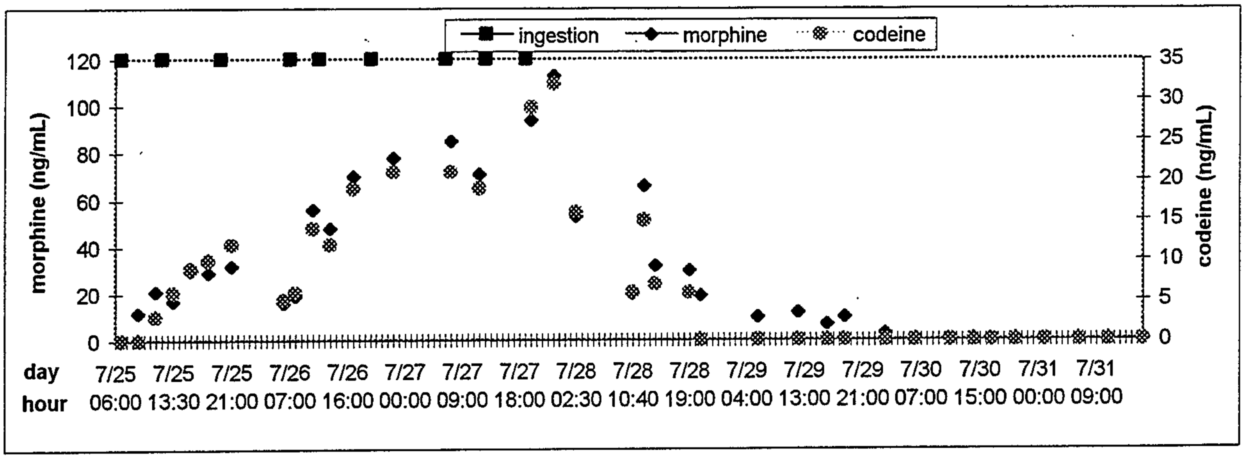


Figure 1. Effect of multiple ingestion of poppy seed on morphine and codeine urinary excretion

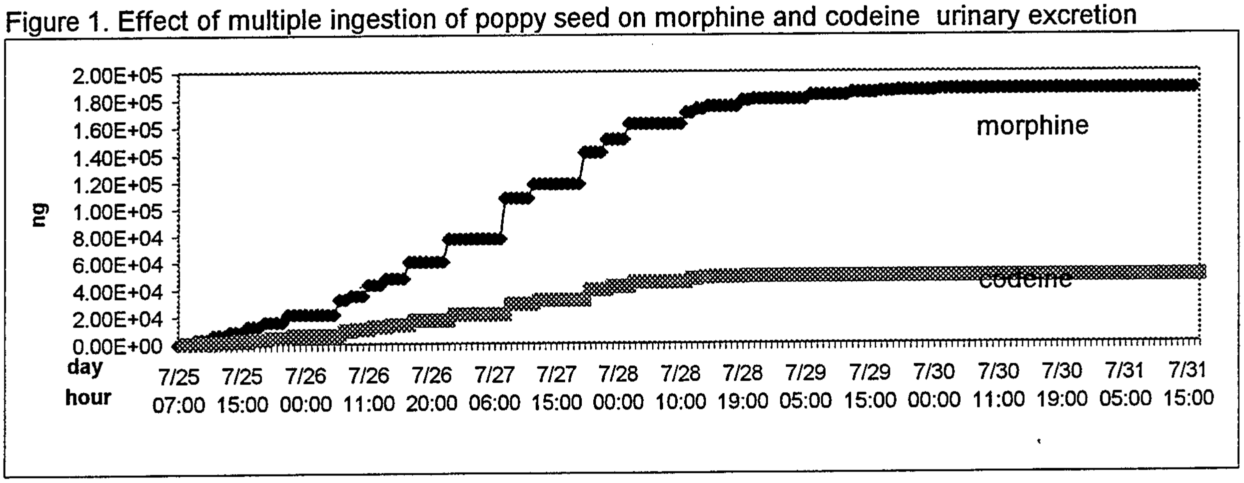


Figure 2. Total urinary excretion of morphine and codeine with multiple ingestion

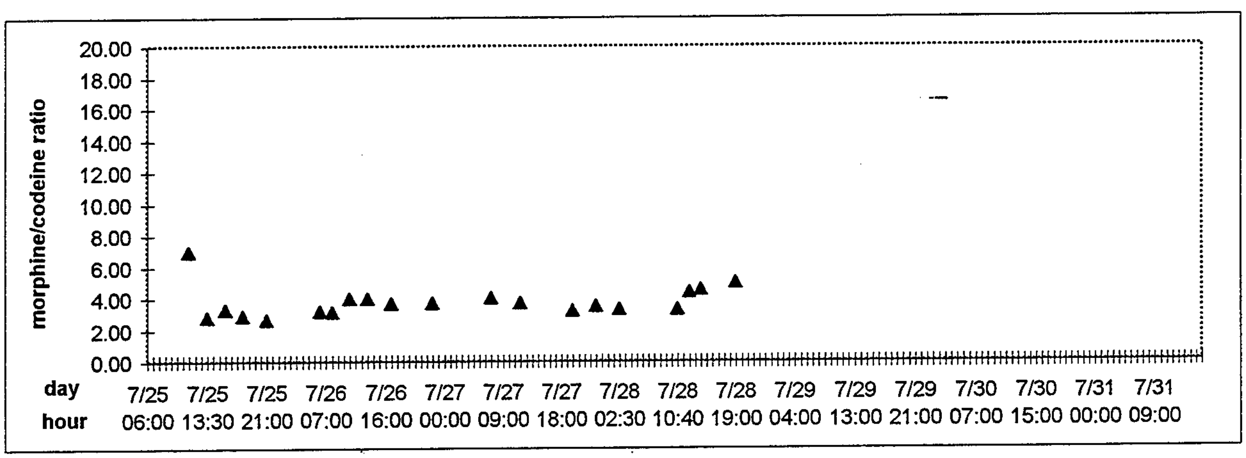


Figure 3. Morphine/codeine ratio with multiple ingestion

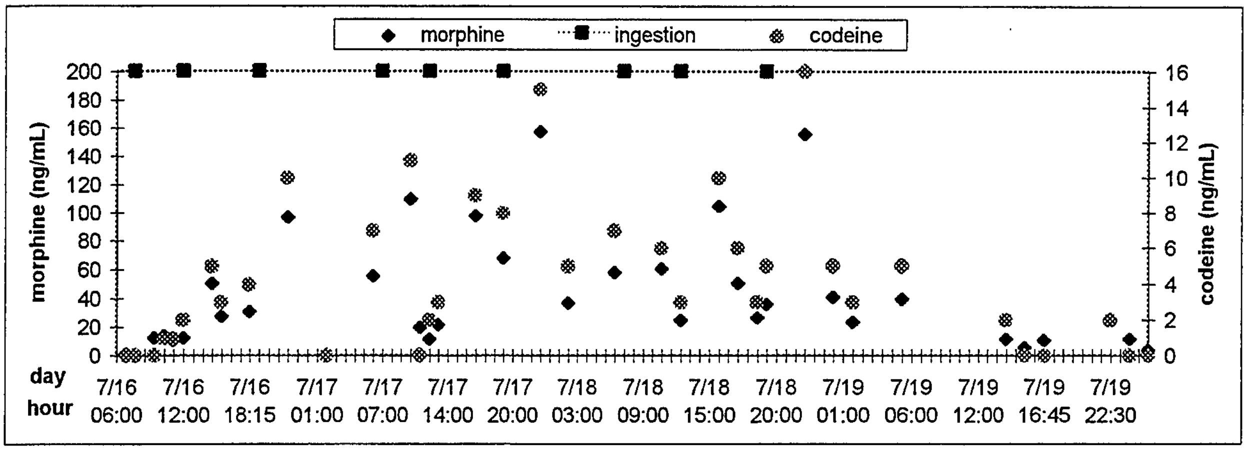


Figure 4. Effect of multiple ingestion of poppy seed on morphine and codeine urinary excretion

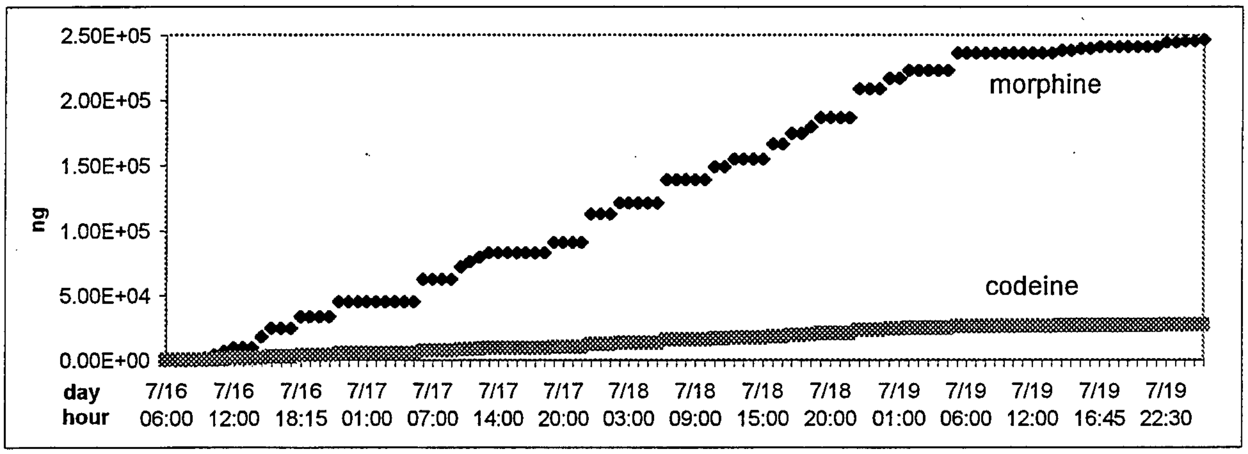


Figure 5. Total urinary excretion of morphine and codeine with multiple ingestion

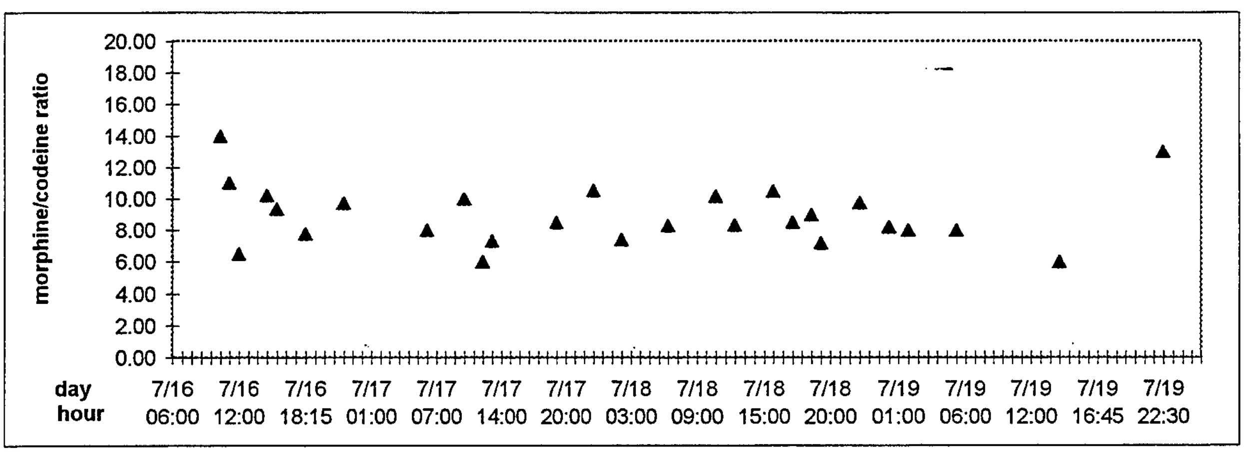


Figure 6. Morphine/codeine ratio with multiple ingestion

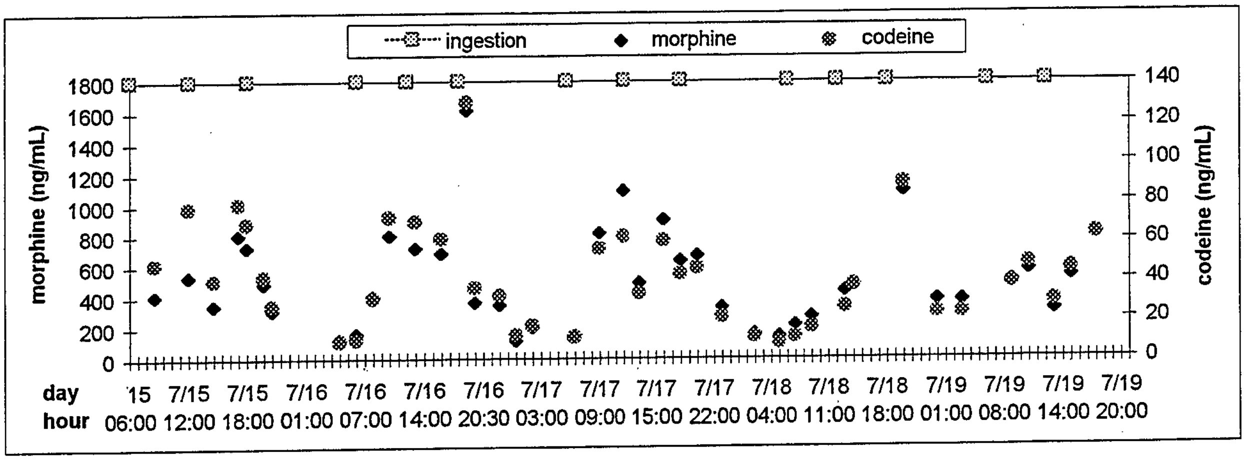


Figure 7. Effect of multiple ingestion of poppy seed on morphine and codeine urinary excretion

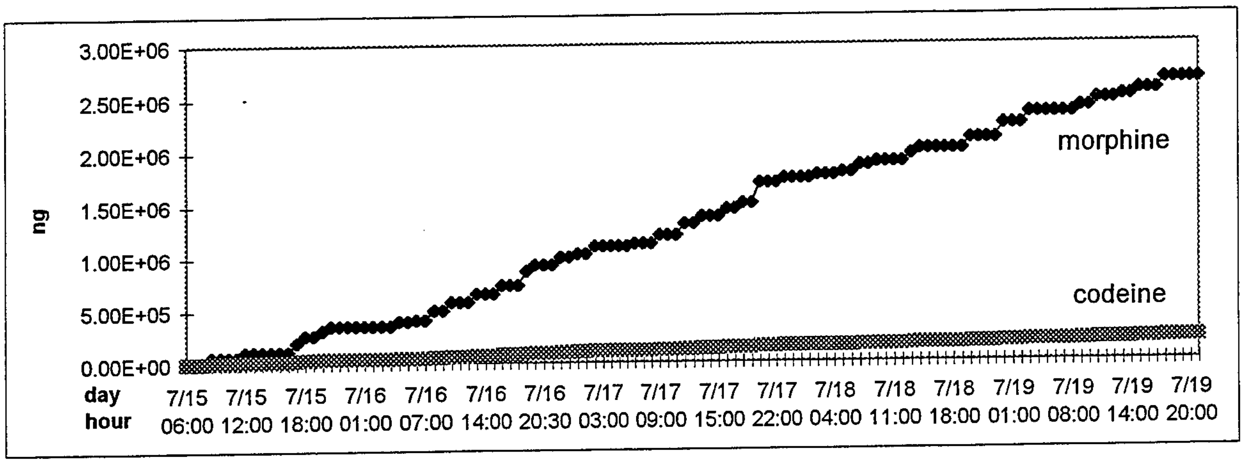


Figure 8. Total urinary excretion of morphine and codeine with multiple ingestion

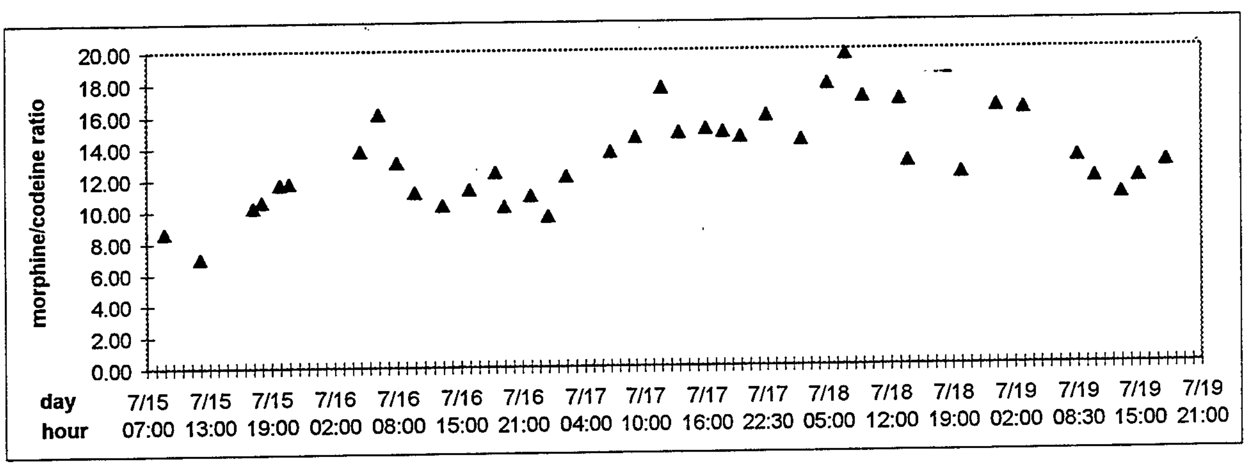


Figure 9. Morphine/codeine ratio with multiple ingestion

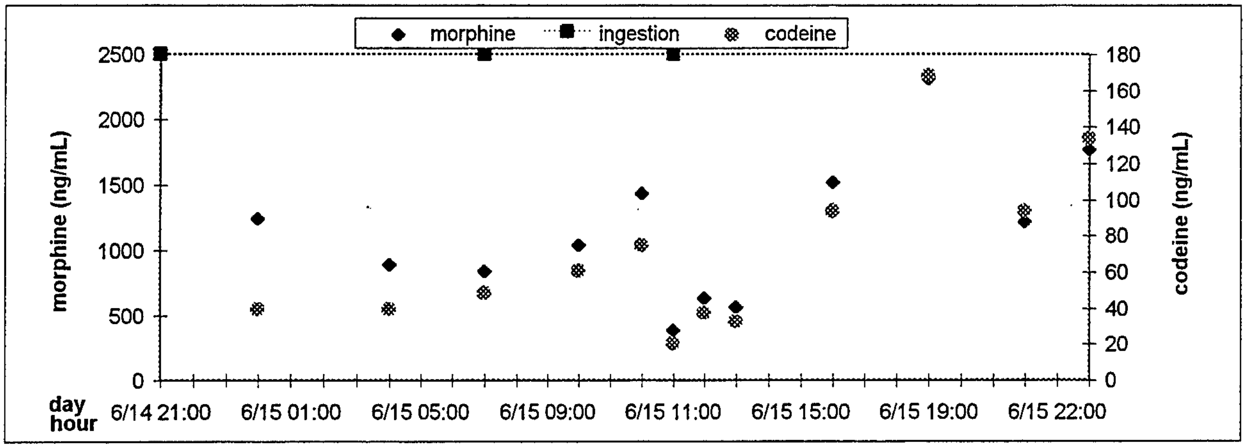


Figure 10. Subject C: Urinary excretion of morphine and codeine following poppy seed ingestion

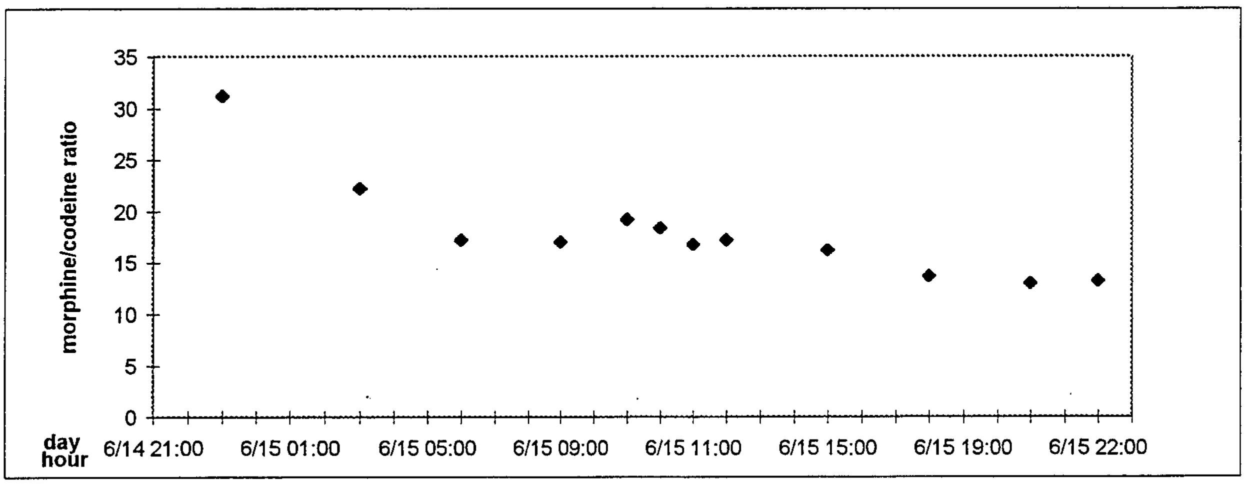


Figure 11. Subject C: Morphine/codeine ratio following poppy seed ingestion

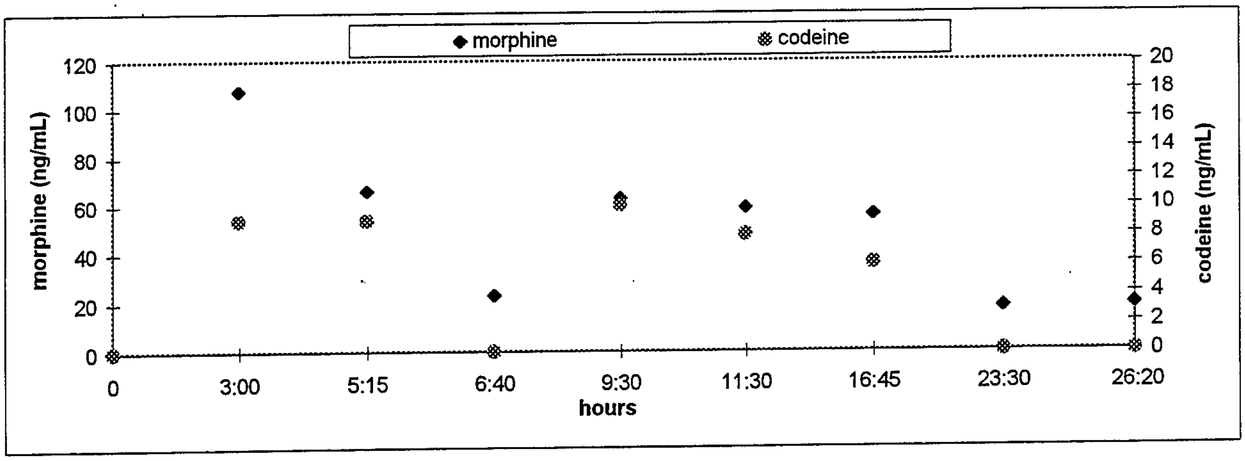


Figure 12. Subject A: Urinary excretion of morphine and codeine following raw poppy seed ingestion

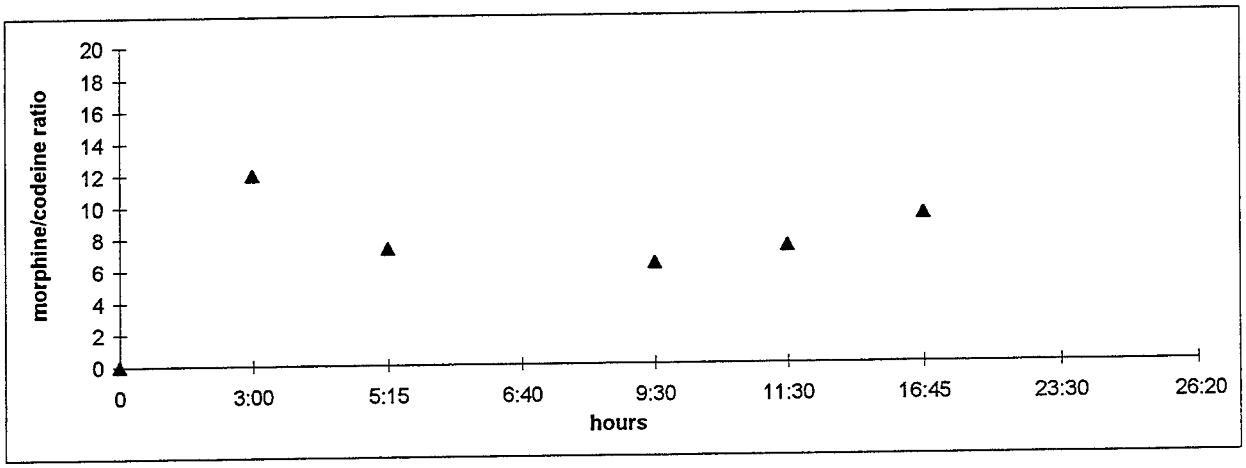


Figure 13. Subject A: Morphine/codeine ratio following raw poppy seed ingestion

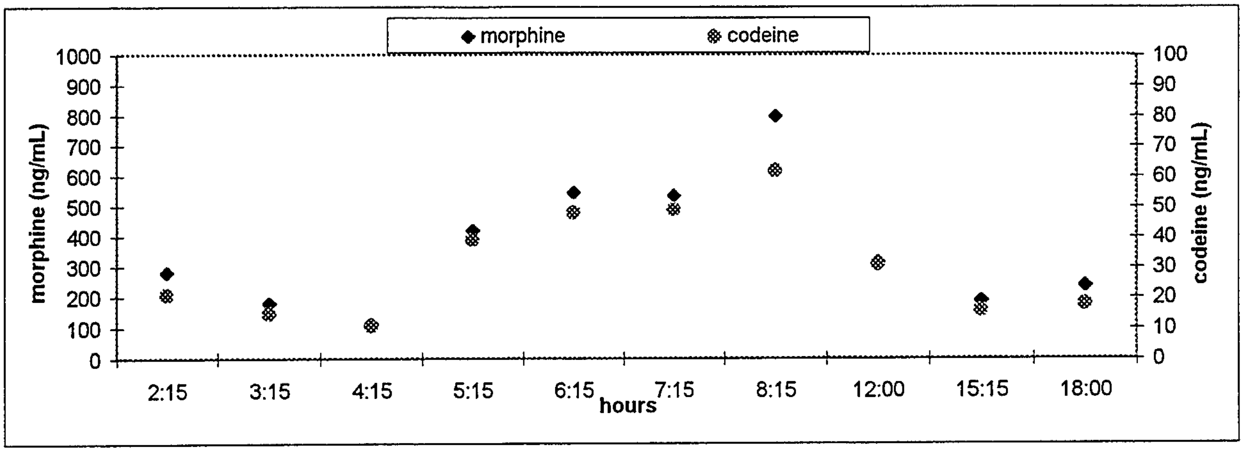


Figure 14. Subject B: Urinary excretion of morphine and codeine following raw poppy seed ingestion

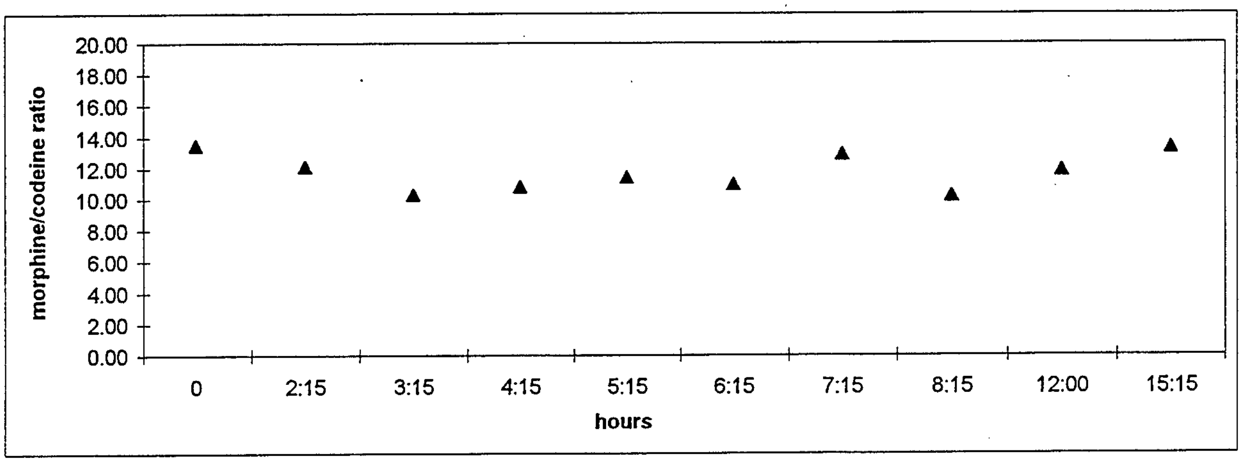


Figure 15. Subject B: Morphine/codeine ratio following raw poppy seed ingestion

References:

Annett, H. E., & Bose, M. N. (1922). The estimation of meconic acid in opium. Analyst, 47, 387-391.

Carpenter, P. A. (1994). Opiates in hair following poppy seed consumption. Thesis, Thomas Jefferson University.

Cone, E. J., Welch, P., Mitchell, J. M., & Paul, B. D. (1991). Forensic drug testing for opiates: 1. Detection of 6-acetylmorphine in urine as an indicator of recent heroin exposure; drug and assay considerations and detection times. J. Analytical Toxicology, 15, 1-7.

Dahlstrom, B. E., & Paalzow, L. K. (1978). Pharmacokinetics interpretation of the enterohepatic recirculation and first-path elimination of morphine in the rat. , 6(6), 505-519.

Duke, J. A. (1985). CRC Handbook of Medicinal Herbs. Boca Raton, Fla.: CRC Press.

ElSohly, H. N., & ElSohly, M. A. (1990). Poppy seed ingestion and opiate urinalysis: A closer look. J. of Analytical Toxicology, 14, 308-310.

ElSohly, H. N., Stanford, D. F., Jones, A. B., ElSohly, M. A., Snyder, H., & Pedersen, C. (1988). Gas Chromatographic/Mass Spectrometric analysis of morphine and codeine in human urine of poppy seed eaters. J. of Forensic Sciences, 33(2), 347-356.

ElSohly, M. A., & Jones.A.B. (1989). Morphine and codeine in biological fluids: approaches to source differentiation. Forensic Sciences Review, 1(1), 13-22.

Fairbairn, J. W., & Steele, M. J. (1981). Meconic acid and alkaloids in Papaver somniferum and P. bracteatum. Planta Medica, 41, 55-60.

Fritschi, G., & Prescott, J. W. R. (1985). Morphine levels in urine subsequent to poppy seed consumption. Forensic Science International, 27, 111-117.

Glare, P. A., & Walsh, T. D. (1991). Clinical pharmacokinetics of morphine. Therapeutic Drug Monitoring, 13, 1-23.

Hayes, L. W., Krasselt, W. G., & Mueggler, P. A. (1987). Concentrations of morphine and codeine in serum and urine after ingestion of poppy seed. Clin. Chem., 33/6, 806-808.

Huestis, M. A., & Cone, E. J. (1995). Drug test findings resulting from unconventional drug exposure. In R. H. Liu & B. A. Goldberger (Eds.), Handbook of Workplace Drug Testing (pp. 289-318). Washington, D.C.: AACC Press.

Lim, H.-Y., & Kwok, S.-F. (1981). Differentiation and comparison of raw, prepared and dross opium. Bulletin on Narcotics, 33, 31-41.

Milne, R. W., Nation, R. L., & Somogyi, A. A. (1996). The disposition of morphine and its 3- and 6-glucuronide metabolites in humans and animals, and the importance of the metabolites to the pharmacological effects of morphine. Drug Metabolism Reviews, 28(3), 345-472.

Miyamoto, S., & Brochmann-Hanssen, E. (1962). Dissociation constants of certain gamma-pyrone dicarboxylic acids. J. Pharmaceutical Sciences, 51, 552-554.

Mule, S. J., & Casella, G. A. (1988). Rendering the "poppy seed defense" defenseless: Identification of 6-monoacetylmorphine in urine by gas chromatography/mass spectroscopy. Clin. Chem., 34(7), 1427-1430.

Pettitt, B. C., Dyazel, S. M., & Hood, L. V. S. (1987). Opiates in poppy seed: Effect on urinalysis results after consumption of poppy seed cake filling. Clin. Chem., 33/7, 1251-1252.

Selavka, C. M. (1991). Poppy seed ingestion as a contributing factor to opiate-positive urinalysis results: the pacific perspective. J. of Forensic Sciences, 36, 685-696.

Struempler, R. E. (1987). Excretion of codeine and morphine following ingestion of poppy seeds. J. of Analytical Toxicology, 11, 97-99.

Vree, T. B., & Verwey Van Wissen, C. P. W. G. M. (1992). Pharmacokinetics and metabolism of codeine in humans. Biopharmaceutics And Drug Disposition, 13, 445-460.

Yamazoe, Y., Numata, H., & Yanagita, T. (1981). Thebaine metabolites in the urine of rhesus monkeys. Japanese J. Pharmacology, 31, 433-442.

Yong, L. H., & Lik, N. T. (1977). The human urinary excretion pattern of morphine and codeine following the consumption of morphine, opium, codeine and heroin. Bulletin on Narcotics, 29(3), 45-73.

APPLICATION OF META-ANALYSIS TO RESEARCH ON PILOT TRAINING: A CASE STUDY OF RESEARCH ON SIMULATOR SCENE CONTENT AND LOW-LEVEL FLIGHT

William A. Stock
Professor
Department of Exercise Science and Physical Education

Arizona State University
Tempe, Arizona 85287-0701

Final Report for:
Summer Faculty Research Program
Armstrong Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, DC

and

Armstrong Laboratory

September 1996

APPLICATION OF META-ANALYSIS TO RESEARCH ON PILOT TRAINING: A CASE STUDY OF RESEARCH ON SIMULATOR SCENE CONTENT AND LOW-LEVEL FLIGHT

William A. Stock
Professor
Department of Exercise Science and Physical Education
Arizona State University

Abstract

This paper reports a meta-analysis conducted of primary research that: (1) examined relations among scene content variables and behaviors relevant to low-altitude flying, and (2) was conducted under the auspices of Armstrong Laboratory. A total of 33 primary research reports were identified using bibliographies available at the Armstrong Laboratory Library. Of these 33, 28 were accessible during the period of this Summer Faculty Research Program. A total of 105 effect sizes were extracted from seven different sources. The average value of these effect sizes was .85; a value that indicates a high degree of positive influence of the manipulations of scene content that were employed in these studies. Based on a global evaluation of this small domain of research studies, a number of recommendations are offered for improving research reporting.

APPLICATION OF META-ANALYSIS TO RESEARCH ON PILOT TRAINING: A CASE STUDY OF RESEARCH ON SIMULATOR SCENE CONTENT AND LOW-LEVEL FLIGHT

William A. Stock

INTRODUCTION

Meta-analysis is a method of quantitatively combining results from primary research reports (Glass, 1976; Cooper and Hedges, 1994). An effective meta-analysis summarizes a domain of research efficiently and provides a basis for making sound policy decisions. The steps necessary to produce an effective meta-analysis include: (1) extracting effect sizes (mean differences or correlations) and other information from primary research reports; (2) coding this information accurately and reliably (Stock, 1994; Stock, Gomez, and Balluerka, 1996), and (3) analyzing the effect sizes in a meaningful manner (Hedges and Olkin, 1985). Here, a constrained meta-analysis was conducted on research on the scene content of simulators that are used to train pilots to fly at low altitudes. The goal of this project was to demonstrate that meta-analysis may be employed to evaluate a domain of research, as well as to summarize the primary results in that domain.

STATEMENT OF THE PROBLEM

For a pilot flying at low altitude, the out-of-the-cockpit, visual scene is complex and changing rapidly. The margins for error are small and the time available for making decisions quite limited. Furthermore, as the contents of the visual scene affect flight and mission decisions made by a pilot, distinguishing critical (and/or sufficient) from noncritical (and/or insufficient) cues in the scene is essential both to the successful completion of a mission and to the survival of the pilot (162nd Tactical Fighter Group, 1986). Therefore, for those who train pilots and for those who design, conduct, and/or evaluate training experiences that occur in simulators, an important goal is to insure that as many critical visual cues as possible are present in the visual scenes of these simulated flights. Unfortunately, the task of determining a minimally sufficient set of visual cues is made more complex by the fact that graphics imaging systems of simulators do not reproduce a real out-of-the-cockpit scene with complete fidelity (Andrews, Carroll, and Bell, 1996). Since the late 1970s and early 1980s (Irish, Grunzke, Gray, and Waters, 1977; Buckland, 1980), there has been a steady and continuing research interest in the effects of scene content on flying behaviors. This research literature was chosen for meta-analysis.

Definition of the Variables of Interest

Chosen as outcome variables of interest were dependent variables that were associated with flying an aircraft at low altitude. This means that pilots or other study participants had to engage in behaviors related to aircraft control (e.g., altitude estimation, stick control). By definition, take offs and landings, and behaviors related to bombing and targeting were not included. Also excluded were

studies in which the simulator controls were fixed at low-altitude settings, and hence, in which participants would not be able to display behaviors related to low-altitude flight.

Scene content variables included such manipulations as: (1) having vertical cues be absent or present, (2) changing the degree of density of objects and/or texture, and (3) creating scenes containing less or more visual detail. These variables are directly related to the fidelity of the visual scene of a simulator.

METHOD

Literature Search for Data Base

The literature search was restricted to studies conducted under the auspices of Armstrong Laboratory. Research studies were initially identified using annotated and categorized bibliographies maintained by the library at Armstrong Laboratory at Williams Gateway Air Field. A total of 33 primary sources were so identified. Of these, 28 were located and used. The remaining five primary sources could not be retrieved in the time available. All 33 sources are listed in Appendix A.

Choice of Effect Size

There are two types of effect sizes. One type is a standardized mean difference on an outcome variable measured at two different levels of a manipulated variable. For example, if an investigator measured the percent of time that pilots maintained their aircraft at a target low altitude both when the scene did and did not contain vertical cues, then an effect size could be computed (given sufficient information is presented in the primary report). The second type of effect size is a correlation between two measured variables. Only effect sizes of the first type were included in the present study. A conceptual formula for the first type of effect size is given by:

$$\text{Effect Size} = \frac{\bar{X}_{\text{First Condition}} - \bar{X}_{\text{Comparison Condition}}}{\text{Std. Deviation}}$$

To apply the above formula, all effect sizes were computed so that the First Condition involved more of the manipulated variable than the Comparison Condition (e.g., more detail, more vertical cues, a greater density of objects, or more texture). Doing so means that positive effect sizes indicate a favorable influence of increasing amounts of the manipulated variable. Further, in repeated measures experiments, the square root of the mean square error for an F-test for an experimental effect was taken as the most appropriate estimate of the standard deviation listed in the above formula (In general, this mean square is some form of a subject by treatment interaction).

To derive an effect size derived from a correlation it would be necessary to manipulate a scene content variable across a set of subjects (each subject receiving a different measured amount of the selected scene content variable) and subsequently measure one or more outcome measures.

These conditions do not ordinarily hold in experimental research. In short, effect sizes derived from correlations are not applicable in this synthesis.

Selection and Coding of Data Base

A meta-analyst attempts to identify attributes of studies that vary with effect sizes. Attributes that are causally linked to effect size magnitude are important because they provide a basis for designing and conducting new empirical research. The information that defines these attributes has to be extracted and coded for analysis. Of a variety of possible categories of items to code (Stock, 1994), a standard set of items would include information about identification of studies, research setting, subjects, methodology, and effect size outcomes. Year and source of publication are examples of identification items. Items in the setting category often describe the use of special populations, as well as the setting in which the study took place. In distinction to general conditions of a study, characteristics of participants of a study are considered subject variables. Items that describe study design and sampling procedures pertain to methodology. Information about effect size forms the final category of items. Included in this category are the summary statistics used to compute effect sizes and information about the outcome measures. The final set of items selected for the present meta-analysis are given in Appendix B.

RESULTS

Table 1 displays how the 33 research sources were sorted at each stage of the synthesis.

Stage	Outcome
Identification of studies	Thirty-three research studies were identified.
Collection of studies	Twenty-eight studies were collected in time.
Assess outcome measure(s)	Twenty-three of the 28 studies had appropriate measure(s).
Assess manipulation of scene content	13 of the 23 studies had experimental manipulation of scene content.
Assess sufficiency of statistics in study.	Nine studies had sufficient statistics to compute an effect size.
Identify number of studies involving pilots.	Eight studies involved pilots.

Of the 28 primary reports, nine involved the use of multidimensional scaling techniques applied to individual samples of subjects and no experimental manipulation. Most of these multidimensional scaling studies involved a minimal use of the simulator. Although 16 of the primary report studies involved the use of the simulator to create, present, or collect responses from subjects, only seven investigations actually involved "flying" the simulator during the course of the investigation. In at least 15 instances, the simulator controls did not function as aircraft controls. Fortunately, 87 of the

105 effect sizes computed were related to conditions in which the subjects were actually flying the simulator.

Figure 1 is a histogram of all 105 effect sizes extracted from this literature data base. An examination of this data base strongly supports the conclusion that manipulating scene content in the direction of more detail, more objects, more vertical cues, and/or more visual and textural detail has a large positive effect -- on the average equivalent to increasing performance .85 standard deviations. Figures 2 and 3 are histograms of the effect sizes associated with introducing objects and increasing the density of cues, respectively. An examination of Figure 2 and 3 supports the same conclusion drawn from Figure 1. The average of eight effect sizes that compared introducing vertical cues (versus no vertical cues) was .95 of a standard deviation unit. These eight effect sizes are not pictured. Relative to typical effect sizes reported in the behavioral science literature, these average effect sizes are quite substantial.

CONCLUSIONS

Overall, effect sizes derived from the nine experimental investigations of scene content demonstrate that manipulations of scene content were reliably related to positive changes in behaviors related to low-altitude flying. The types of changes in scene content that led to improved performance included introduction of objects and texture, the introduction of vertical stimuli like trees, hills (even inverted tetrahedrons), and increases in the density of objects in the visual scene. This research confirms what pilots had been telling researchers since the initial research studies on scene content (Buckland, 1980). Nevertheless, the magnitude of the average effect size clearly documents that most changes in scene content that create more realistic visual scenes dramatically improve the performance of behaviors related to low-level flying.

A disappointing outcome of this project was the inability to extract effect sizes from a number of research reports. The inability to extract effect sizes stems primarily from a lack of sufficient statistics being reported in primary research reports. Fortunately, this is a technical problem that can be addressed by simple changes in the content of the research reports created under the auspices of Armstrong Laboratory. At a minimum, research investigators should include the sufficient statistics associated with their studies in their reports. For example, in multiple-group, multivariate, studies authors need to report a pooled within-group, variance-covariance matrices, and vectors of means and sample sizes. In univariate research, the basic notion is to report summary tables of means and standard deviations so that later researchers can compute effect sizes on any experimental comparison possible.

Another interesting outcome of this project was the finding that a total of 9 different primary research reports involved the application of multidimensional scaling techniques. In retrospect, it is difficult for this investigator to construct a rationale for conducting so many investigations in this domain employing this technique. There are three problems associated with the application of the

scaling methodology in this domain of training research. First, no outcome variables are related to the behavior of primary interest: successfully flying at low altitude. Second, there are no experimental comparisons. Third, results from multidimensional scaling studies are highly dependent on the choice of stimuli employed in the investigation.

Finally, it should be noted that this project would not have been possible without the commitment of behavior scientists of Armstrong Laboratory to document their research efforts, and the efforts of the Armstrong Laboratory librarian to maintain a complete and accessible inventory of past research.

REFERENCES

162nd Tactical Fighter Group. (1986). *Academic text: Low-altitude training.* Tucson, AZ: Author.

Andrews, D. H., Carroll, L. A., & Bell, H. H. (1995). *The future of selective fidelity in training devices.* (AI/HR-TR-1995-0195). Aircrew Training Research Division, 6001 South Power Road, Mesa, AZ.

Buckland, G. H. (1980). Visual cue requirements for terrain flight simulation. In *Proceedings of the 2nd Annual Interservice/Training Equipment Conference and Exhibition* (pp. 92-93). Arlington, VA: National Security Industrial Association.

Casey, E. P., & McConnon, M, E. (1995). *Annotated bibliography of Aircrew Training Research Division technical publications.* (AI/HR-TR-1995-0079). Aircrew Training Research Division, 6001 South Power Road, Mesa, AZ.

Cooper, H., & Hedges, L. V. (Eds.). (1994). *Handbook of research synthesis.* New York: Russell Sage.

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher, 5,* 3-8.

Hedges, L. V., & Olkin, I. (1985).*Statistical methods for meta-analysis.* Orlando, FL: Academic Press.

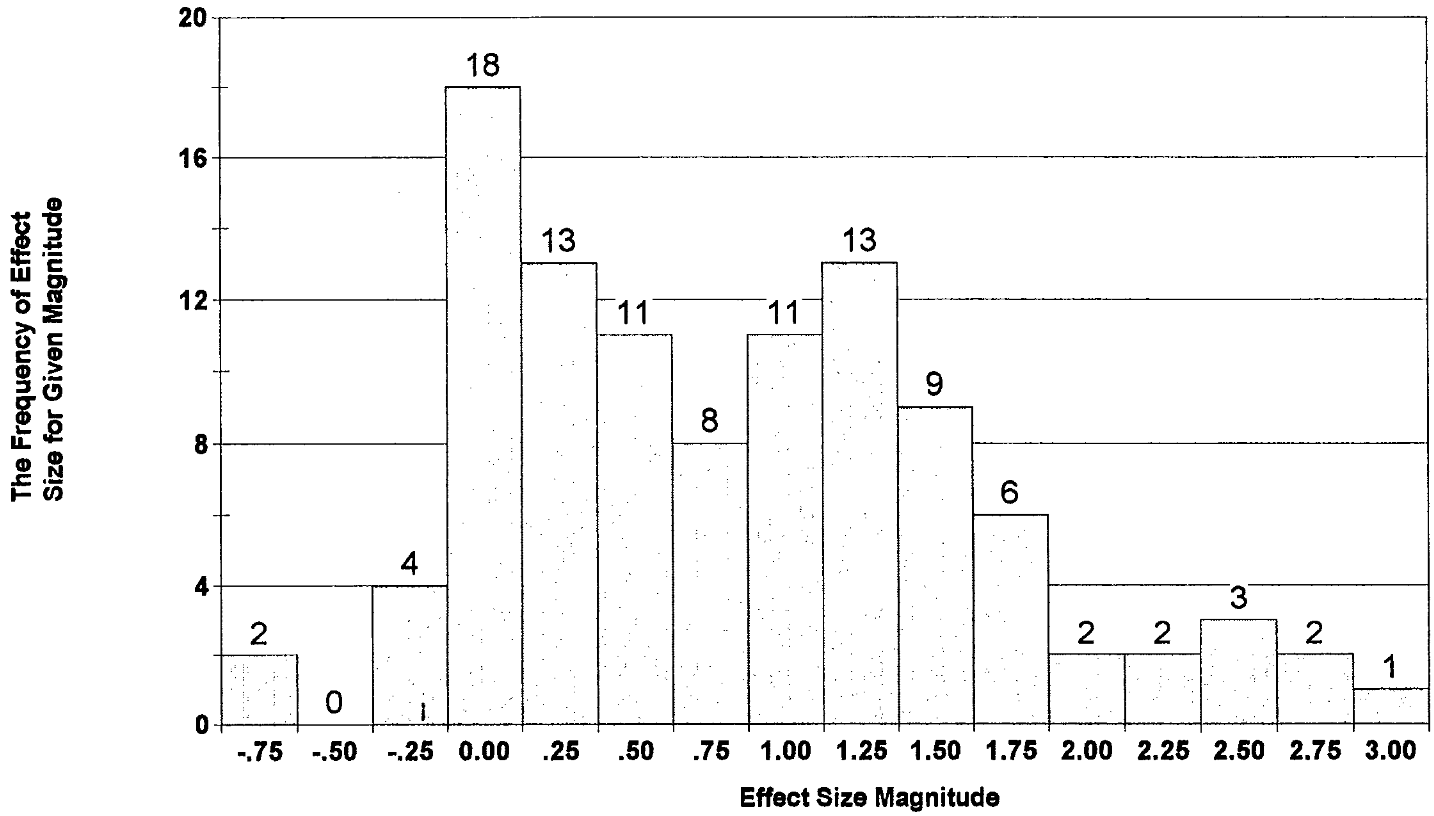
Irish, P. A., III, Grunzke, P. M., Gray, T. H., & Waters, B. K. (1977). *The effects of system and environmental factors upon experienced pilot performance in the advanced simulator for pilot training.* (AfHRL-TR-77-13). Williams AFB, AZ: Flying Training Division, Air Force Human Resources Laboratory.

Stock, W. A. (1994). Systematic coding for research synthesis. In H. Cooper, & L. V. Hedges (Eds.), *Handbook of research synthesis.* (pp. 139-162). New York: Russell Sage Foundation.

Stock, W. A., Gómez Benito, J., & Balluerka Lasa, N. (1996). Research synthesis: Coding and conjectures. *Evaluation and the Health Professions, 19*, 104-117.

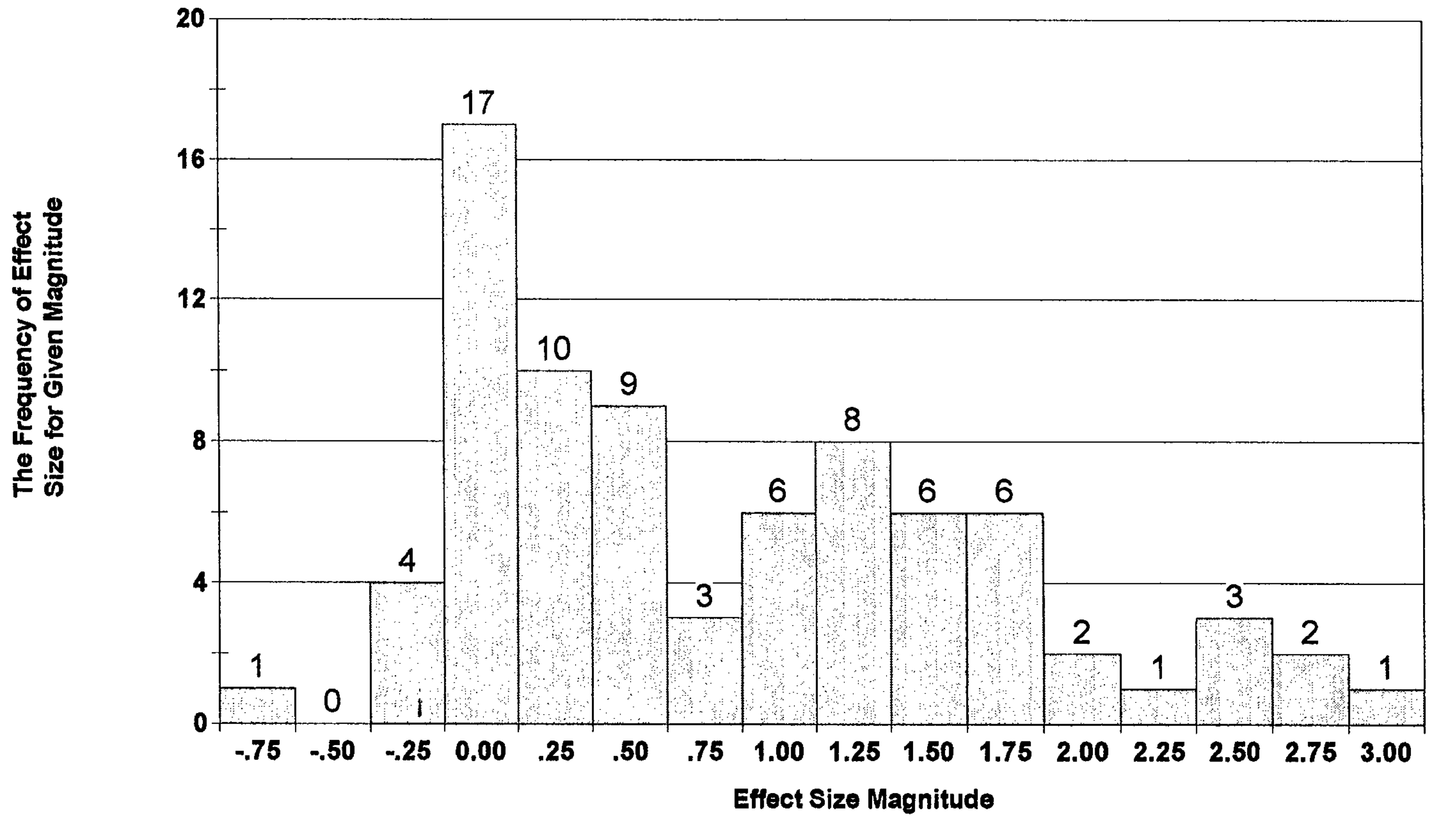
Warner, H. D., & Casey, E. P. (1995). *Flight simulator visual system research and development: Bibliography of support provided by the Aircrew Training Research Division.* (AI/HR-TR-1995-0107). Aircrew Training Research Division, 6001 South Power Road, Mesa, AZ.

Figure 1. A Frequency Histogram of All Effect Sizes.



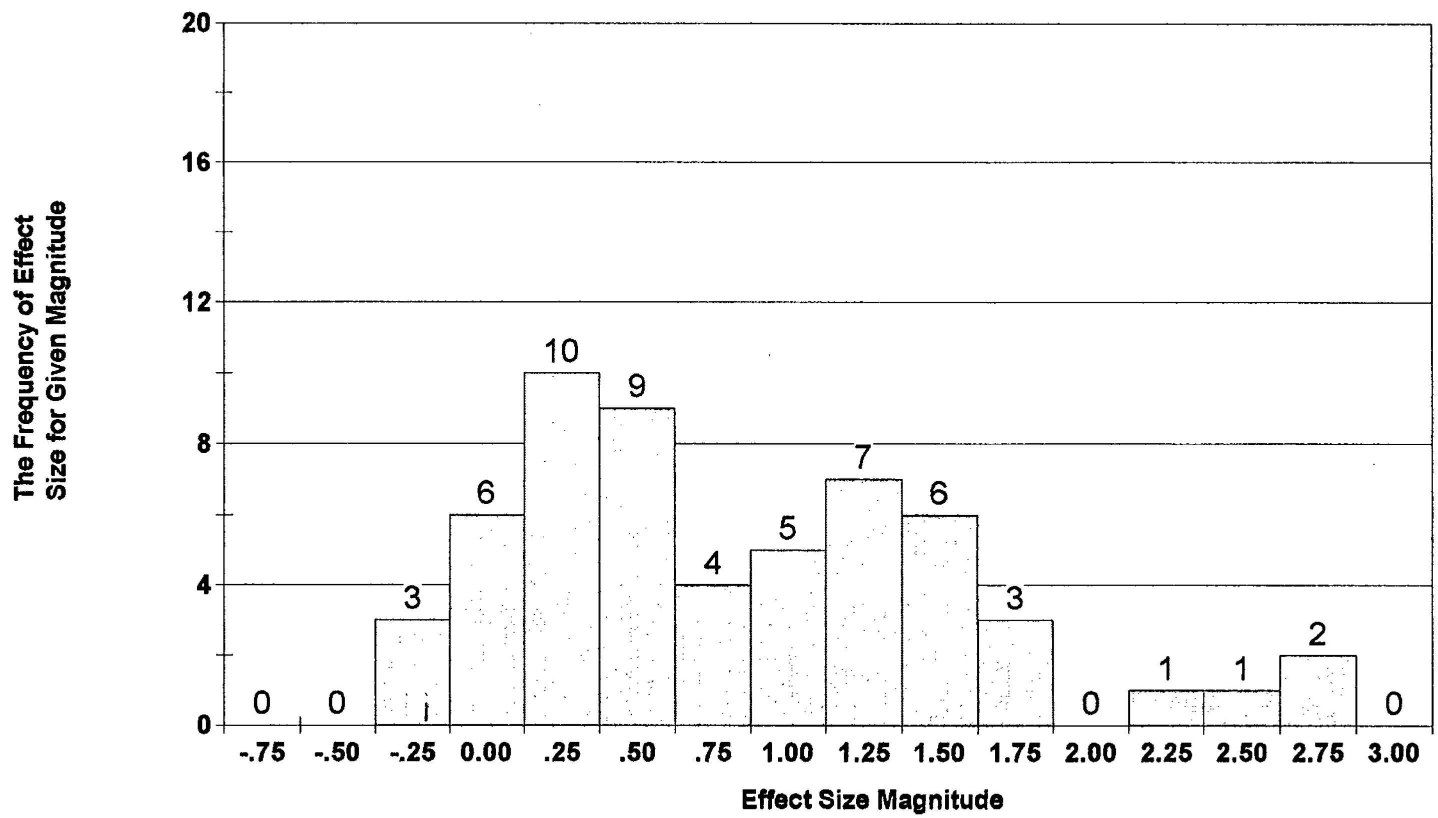
Note. The Mean and Standard Deviation are .84 and .80, respectively.

Figure 2. A Frequency Histogram of All Effect Sizes Involving Object Comparisons.



Note. The Mean and Standard Deviation are .82 and .87, respectively.

Figure 3. A Frequency Histogram of All Effect Sizes Involving Density Comparisons.



Note. The Mean and Standard Deviation are .83 and .74, respectively.

Appendix A: Initial Literature Data Base

Bibliography of Reports Initially Judged Eligible for Inclusion in the Research Synthesis

Buckland, G. H. (1980). Visual cue requirements for terrain flight simulation. In Proceedings of the 2nd Intersevice/Training Equipment Conference and Exhibition (pp. 92-93). Arlington, VA: National Security Industrial Association.

Buckland, G. H., Edwards, B. J., & Stephens, C. W. (1981). Flight simulator visual and instructional features for terrain flight simulation. In E.G. Monroe (Ed.), Proceedings of the 1981 IMAGE II Conference (AFHRL-TR-81-48, AD-A110 226, pp. 350-362). Williams AFB, AZ: Operations Training Division, Air Force Human Resources Laboratory.

DeMaio, J., Bell, H. H., & Brunderman, J. (1985). Pilot-oriented performance measurement (AFHRL-TP-85-18, AD-A158 849). Williams AFB, AZ: Operations Training Division, Air Force Human Resources Laboratory

DeMaio, J., & Brooks, R. (1982). Assessment of simulator visual cueing effectiveness by psychophysical methods. In Proceedings of the 4th Intersevice/Training Equipment Conference (Vol. 1, pp. 379-381). Arlington, VA: National Security Industrial Association.

DeMaio, J., Brooks, R., Brunderman, J., & Rinalducci, E. J. (1983). Visual cueing effectiveness – Comparison of perception and flying performance. In A. T. Pope & L. D. Haugh (Eds.) Proceedings of the Human Factors 27th Annual Meeting (Vol. 2, pp. 929-932). Santa Monica, CA: Human Factors Society.

DeMaio, J., Rinalducci, E. J., Brooks, R., & Brunderman, J., (1983). Visual cueing effectiveness: Comparison of perception and flying performance. . In Proceedings of the 5th Intersevice/Training Equipment Conference (Vol. 1, pp. 92-96). Arlington, VA: American Defense Preparedness Association..

Dixon, K. W., & Curry, D. G. (1987). Effect of scene content and field of view on weapons delivery training. In Proceedings of the 9th Intersevice/Training Equipment Conference (pp. 247-256). Arlington, VA: American Defense Preparedness Association..

Dixon, K. W., & Curry, D. G. (1990). Weapons delivery training: Effects of scene content and field of view. (AFHRL-TP-88-29, AD-A227 968). Williams AFB, AZ: Operations Training Division, Air Force Human Resources Laboratory.

Irish, P. A., III, & Buckland, G. H. (1978). Effects of platform motion, visual and G-seat factors upon experienced pilot performance in the flight simulator. (AFHRL-TR-78-9, AD-A055 691). Williams AFB, AZ: Flying Training Division, Air Force Human Resources Laboratory.

Irish, P. A., III, Grunzke, P. M., Gray, T. H., & Waters, B. K. (1977). The effects of system and environmental factors upon experienced pilot performance in the advanced simulator for pilot training. (AFHRL-TR-77-13). Williams AFB, AZ: Flying Training Division, Air Force Human Resources Laboratory.

Kleiss, J. A. (1990). Terrain visual cue analysis for simulating low-level flight: A multidimensional scaling approach. (AFHRL-TR-90-20, AD-A223 564). Williams AFB, AZ: Operations Training Division, Air Force Human Resources Laboratory.

Kleiss, J. A. (1990). Terrain visual cue analysis for simulating low-level flight: A multidimensional scaling approach. IN E. G. Monroe (Ed.) Proceedings of the 1990 IMAGE V Conference (pp.60-67). Tempe, AZ: IMAGE Society, Inc.

Kleiss, J. A. (1991). Cueing and scene content requirements for low-level flight. In Visual Issues in Training and Simulation Presentation Summaries (pp.31-32). Williams AFB, AZ: Aircrew Training Research Division, Armstrong Laboratory.

Kleiss, J. A. (1991). Multidimensional scaling analysis of terrain features relevant for simulating low-level flight. In Proceedings of the Human Factors Society 35th Annual Meeting (Vol. 2, pp.1372-1376). Santa Monica, CA: Human Factors Society.

Kleiss, J. A. (1992). Influence of operational factors on importance of scene properties for visual low-altitude flight. (AL-TR-1992-0158). Williams AFB, AZ: Aircrew Training Research Division, Armstrong Laboratory.

Kleiss, J. A. (1992). Perceptual dimensions of visual scenes relevant for simulating low-altitude flight. (AL-TR-1992-0011 AD-A254 645). Williams AFB, AZ: Aircrew Training Research Division, Armstrong Laboratory.

Kleiss, J. A. (1992). Tradeoffs among types of scene detail for simulating low-altitude flight. In Proceedings of the 1992 IEEE International Conference on Systems, Man, and Cybernetics (Vol. 2, pp.1121-1146). New York, NY: Institute of Electrical and Electronics Engineers.

Kleiss, J. A. (1993). Properties of computer-generated scenes important for simulating low altitude flight. In Proceedings of the Human Factors and Ergonomics Society 37th Annual Meeting (Vol. 1, pp.98-102). Santa Monica, CA: Human Factors and Ergonomics Society.

Kleiss, J. A. (1994). Effect of terrain shape and object grouping on detection of altitude change in a flight simulator. In Proceedings of the Human Factors and Ergonomics Society 38th Annual Meeting (Vol. 1, pp.119-122). Santa Monica, CA: Human Factors and Ergonomics Society.

Kleiss, J. A. (1994). Perceptual dimensions of simulated scenes relevant for visual low-altitude flight. (AL/HR-TR-1994-0170). Williams AFB, AZ: Aircrew Training Research Division, Armstrong Laboratory.

Kleiss, J. A. (1995). Information perceived in simulated scenes displayed with reduced field of view: A further experiment. (AL/HR-TR-1995-0016). Williams AFB, AZ: Aircrew Training Research Division, Armstrong Laboratory.

Kleiss, J. A. (1995). Identification of environmental elements important for simulating visual low-altitude flight. (AL/HR-TR-1995-0080). Williams AFB, AZ: Aircrew Training Research Division, Armstrong Laboratory.

Kleiss, J. A., Curry, D. G., & Hubbard, D. C. (1988). Effect of three-dimensional object type and density in simulated low-level flight. In Proceedings of the Human Factors Society 32nd Annual Meeting (Vol. 2, pp.1299-1303). Santa Monica, CA: Human Factors and Ergonomics Society.

Kleiss, J. A., Hubbard, D. C., & Curry, D. G. (1989). Effect of three-dimensional object type and density in simulated low-level flight. (AFHRL-TR-88-66, AD-A209 756).

Williams AFB, AZ: Operations Training Research Division, Air Force Human Resources Laboratory.

Kleiss, J. A., & Hubbard, D. C. (1993). Effect of two types of scene detail on detection of altitude change in a flight simulator. (AL-TR-1991-0043, AD-A242 034). Williams AFB, AZ: Aircrew Training Research Division, Armstrong Laboratory.

Kleiss, J. A., & Hubbard, D. C. (1991). Effect of three types of flight simulator visual scene detail on detection of altitude change. Human Factors, 35(4), 653-671.

Martin, E. L., & Rinalducci, E. J. (1983). Low-level flight simulation: Vertical cues (AFHRL-TR-83-17, AD-A133 612). Williams AFB, AZ: Operations Training Research Division, Air Force Human Resources Laboratory.

McCormick, D., Smith, T., Lewandowski, F., Preskar, W., & Martin, E. (1983). Low-Altitude Database Development Evaluation and Research (LADDER). In Proceedings of the 5th Interservice/Industry Training Equipment Conference (Vol. 1, pp. 150-155). Arlington, VA: American Defense Preparedness Association.

Rinalducci, E. J., De Maio, J., Patterson, M. J., & Brooks, R. (1983). Psychophysical assessment of simulator visual displays. In R. S. Jensen (Ed.), Proceedings of the Second Symposium on Aviation Psychology (pp. 489-494). Columbus, OH: Ohio State University.

Rinalducci, E. J., Martin, E. L., & Longridge, T. (1982). Visual cues in the simulation of low-level flight. In J. F. Swiney (Ed.), Proceedings of the Eighth Symposium on Psychology in the Department of Defense (USAFA-TR-82-10, pp. 32-36). Colorado Springs, CO: Department of Behavioral Sciences and Leadership, U. S. Air Force Academy.

Rinalducci, E. J., Patterson, M. J., & De Maio, J. (1984). Static vs. Dynamic presentation of visual cues in simulated low-level flight. In Proceedings of the Ninth Symposium on Psychology in the Department of Defense (USAFA-TR-84-2, AD-A141 043, pp. 667-671). Colorado Springs, CO: Department of Behavioral Sciences and Leadership, U. S. Air Force Academy.

Warner, H. D., Serfoss, G. L., & Hubbard, D. C. (1993). Effects of area-of-interest display characteristics on visual search performance and head movements in simulated low-level flight. (AL-TR-93-0023). Williams AFB, AZ: Aircrew Training Research Division, Air Force Human Resources Laboratory.

Warner, H. D., Serfoss, G. L., & Hubbard, D. C. (1994). Altitude cueing effectiveness of terrain texture characteristics in simulated low-level flight. (AL-TR-94-0168). Williams AFB, AZ: Aircrew Training Research Division, Air Force Human Resources Laboratory.

Appendix B: Meta-analysis Coding Forms

Study Characteristics

Synthesis ID No.		Technical Report Number			Year Published	
If not a Technical Report, check the type of publication						
	Article		Paper/Proceedings		Edited Vol., Book		Other
List all authors in order						
Did this paper/study relate scene content to in low altitude flying (exclusive of takeoff and landing performance)?					Yes	
					No	
Did this paper/study report sufficient statistics to estimate one or more effect sizes (either experimental or correlational)?					Yes	
					No	
Did this paper/study report results of more than one experiment?					Yes	
					No	
How many experiments/studies were reported?						
Comments:						

Global Sample Characteristics

Total number of subjects		
Total Number of pilots		
For the pilots, check the types of aircraft in which they have experience. Check all that apply.	F-15	
	F-16	
	A-10	
	Student Pilot in Fighter or other training.	
	OTHER (List):	
Mean flying hours for these pilots?		
If some subjects were not pilots, briefly characterize the type of subjects used.		

Synthesis ID No.____

Methodological Characteristics of the Study

Location of the Study?	Location. Code:	
Was this a true experiment?	Yes	
	No	
If a true experiment, describe the nature of the study.		
	Experimental coding 1	
	Experimental coding 2	
If not a true experiment, describe the nature of the study.		
	Nonexperimental coding 1	
	Non experimental coding 2	
Was a plausible alternate methodology available?	Yes	
	No	
Describe the nature of the alternate methodology.		

Simulator Use in the Study

Was a simulator used in any way for any part of the experiments/studies reported?	Yes	
	No	
If a simulator was used, describe the nature of the use.		
	Simulator use coding 1	
	Simulator Use coding 2	
If app., total # simulator hours?		
If app., total # simulator sessions?		
If app., type of simulator used?		
If app., location of simulators)?		
Comments:		

Synthesis ID No. _____

Subsample Characteristics

Subsample Description and Number		
Was this an Experimental Group	Yes	
	No	
Were there pilots in this group	Yes	
	No	
How many were pilots?		
Flying experience	Aircraft	Hours
Describe any other salient pilot attribute.		

SFRP: Research Syn.: Page 4

Synthesis ID No. ______

Effect Size Information

Experimental Effect Size Information										
Scene content description	Outcome Measure	Group No.	Mean	SD	N	Scene content description	Group No.	Mean	SD	N

Correlation Effect Sizes (If any)			
Group	Measure 1/Measure 2 descriptions	R	N

Other types of effect or quasi-effect sizes and/or Global Methodological Conclusion

ENGAGEMENT, INVOLVEMENT, AND SELF-REGULATED LEARNING: CONSTRUCT AND MEASUREMENT DEVELOPMENT TO ASSESS ACHIEVEMENT AND CALIBRATION

Nancy J. Stone
Assistant Professor
Department of Psychology

Creighton University
2400 California Plaza
Omaha, NE 68178

Final Report for:
Summer Faculty Research Program
Armstrong Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, DC

and

Armstrong Laboratory

September, 1996

ENGAGEMENT, INVOLVEMENT, AND SELF-REGULATED LEARNING: CONSTRUCT AND MEASUREMENT DEVELOPMENT TO ASSESS ACHIEVEMENT AND CALIBRATION

Nancy J. Stone
Assistant Professor
Department of Psychology
Creighton University

Abstract

In order to determine ways to develop better evaluation of student learning and to design better educational tools to enhance student learning, the areas of student engagement, involvement, and self-regulated learning were thoroughly reviewed. It was discovered that engagement, involvement, and self-regulated learning were related, yet somewhat different areas of research related to student learning. Because these research areas had some overlap in terminology, the constructs of engagement, involvement, and self-regulated learning were clarified. Furthermore, it was determined that these three constructs were related to student achievement. Although untested, it is likely that there is also a relationship between these three constructs and calibration. In order to test these relationships, though, a reliable and valid measure is required. Unfortunately, most researchers tend to develop their own scales which are specific to their research project. This implies that there are no common measures and construct validity is limited. Hence, measurement items for engagement, involvement, and self-regulation were developed. Finally, research was proposed to test the reliability and validity of these measures, to determine the relationship between these measures and student achievement and calibration, and to evaluate ways to increase students' levels of engagement, involvement, and self-regulated learning. Knowledge gained from the proposed measures and research will be extremely beneficial in determining which students possess the skill or attitude to succeed, what teaching or training designs are needed to enhance student engagement, involvement, and/or self-regulated learning, and how to increase student calibration.

ENGAGEMENT, INVOLVEMENT, AND SELF-REGULATED LEARNING: CONSTRUCT AND MEASUREMENT DEVELOPMENT TO ASSESS ACHIEVEMENT AND CALIBRATION

Nancy J. Stone

In educational and training environments, it is critical that students and trainees have the opportunity to assimilate and master as much of the information which is presented to them in order to increase achievement. Additionally, it would be greatly beneficial if students knew what they did and did not know with some accuracy. That is, students should be well calibrated. Interestingly, engagement (e.g., Finn, Folger, & Cox, 1991; Skinner & Belmont, 1993), involvement (e.g., Reed & Schallert, 1993), and self-regulated learning (e.g., Zimmerman, 1986, 1990) appear to be three related, yet somewhat distinct areas of research which address the process by which students do or do not acquire knowledge. Because college and many training environments are much less controlled than elementary and/or secondary schools (specifically concerning study time), it is important to identify what determines whether individuals will become totally engrossed in, own, or be responsible for their learning process.

From the literature, and based mostly on teacher observations, engagement may be considered to span along a continuum from disengaged to engaged. Disengaged elementary students tend to display restless behavior, to annoy others, and to need reprimands (Finn et al., 1991). Additionally, to be disengaged is to be passive, to expend little effort, to give up easily, and to be bored, depressed, anxious, angry, withdrawn from learning, and rebellious (Skinner & Belmont, 1993). At the other extreme, fully engaged students tend to focus on achieving a deep understanding of the material (Ainley, 1993), to display initiative (Finn et al., 1991; Lee & Anderson, 1993; Skinner & Belmont, 1993), to study/work beyond (course) requirements (Finn et al., 1991; Lee & Anderson), to be thorough (Finn et al., 1991), to verbally discuss ideas with others (Finn et al., 1991; Goff & Ackerman, 1992), to be completely absorbed in one's work whereby time becomes distorted (Goff & Ackerman, 1992), to exhibit intense concentration (Helstrup, 1989; Skinner & Belmont, 1993), to desire engagement (Goff & Ackerman, 1992), to display persistence and effort (Finn et al., 1991; Skinner & Belmont, 1993), to challenge their abilities (Skinner & Belmont, 1993), and to display positive affect (Skinner & Belmont, 1993).

Interestingly, students identified as "involved" have also been described as absorbed in their work, displaying intense concentration, challenging their abilities, and exhibiting positive

affect (Reed & Schallert, 1993). Additionally, engaged students have been described as involved in class activities (Lee & Anderson, 1993). Even though the concepts of engagement and involvement appear to be similar, it is proposed that engagement refers to intense concentration on, attention to, and absorption in a task as well as a desire to thoroughly learn the material and to learn more beyond the specified requirements. Because involvement has been measured using adapted job and work involvement scales (Farrell & Mudrack, 1992), involvement is hypothesized to tap the learner's commitment to, perceived importance of, and ownership of learning.

Self-regulated learning may also be distinguished from engagement and involvement, although self-regulated learning and engagement appear to share some common elements. From this review of the literature, the only apparent overlap between engagement and self-regulated learning occurred when terminology suggested that students displayed initiative (Schunk, 1990), a desire to engage (McCombs & Marzano, 1990), persistence (Pintrich & de Groot, 1990; Schunk, 1990; Zimmerman, 1986) and effort (Pintrich & de Groot, 1990). In contrast, self-regulated learners were described as autonomous (Dickinson, 1992; McCombs & Whisler, 1989), self-regulated or self-monitored (Corno & Mandinach, 1983; McCombs & Marzano, 1990; McCombs & Whisler, 1989; Pintrich & de Groot, 1990), resourceful (Zimmerman, 1990), and strategic (McCombs & Whisler, 1989). Figure 1 presents the relationship between the qualities of engaged, involved, and self-regulated learners.

According to Butler and Winne's (1995) model, a self-regulated learner will set goals based on an evaluation of a given task, use strategies to meet those goals, and monitor activities to assess progress and to make a reinterpretation of the task, given feedback (internal or external) about the task. Self-regulated learning may be conceived as autonomous learning (Dickinson, 1992; McCombs & Whisler, 1989), which includes self-evaluation (Butler & Winne, 1995; Zimmerman, 1990), self-regulation, or monitoring (Butler & Winne, 1995; Corno & Mandinach, 1983; McCombs & Marzano, 1990; McCombs & Whisler, 1989; Pintrich & de Groot, 1990), resulting in feedback (Butler & Winne, 1995; Zimmerman, 1990) used for re-assessment of the task. Self-regulated learning is posited to reflect a more systematic process by which students attain a learning goal through the gathering (acquisition) and manipulation (categorizing, organizing, processing) of information, strategic planning (setting goals, monitoring progress), and feedback to continue the self-evaluative cycle.

Fortunately, performance or achievement increases when students are engaged (Ainley, 1993; Goff & Ackerman, 1992; Greenwood, Terry, Marquis, & Walker, 1994; Helstrup, 1989; Reed & Schallert, 1993; Thomas, Bol, Warkentin, Strage, & Rohwer, 1993), involved (Farrell & Mudrack, 1992; Jacobi, 1991), or self-regulated (Butler & Winne, 1995; Pintrich & de Groot, 1990). Increases in performance are most likely due to the strategies associated with engagement, involvement, and self-regulated learning. Although engagement may be described and defined, identified strategies are less common. Yet, if someone were to think abstractly, to consider the benefits of the thought process employed, or to try to create new solutions (Goff, cited in Ackerman & Goff, 1994), this person may become more engaged, but it does not guarantee engagement. This, of course, still remains to be empirically studied. The only specifically identified strategy of engagement found in the literature was the development of associations (Helstrup, 1989).

No strategies of involvement were found, but several self-regulated learning strategies have been identified. Self-regulated learning strategies also include the development of associations (McCombs & Whisler, 1989) via elaboration (Corno & Mandinach, 1983; Pintrich & de Groot, 1990), by integrating material, or by deciding which material is relevant or irrelevant to the topic (Corno & Mandinach, 1983). The difference between engagement and self-regulated learning associations may be that engaged students continuously make connections among the new and old information, which may take them off task. Self-regulated learners, though, would monitor the effectiveness of the associations in relation to the goal.

Besides elaboration, self-regulated learners employ rehearsal (Pintrich & de Groot, 1990), generative (construction of summaries) or duplicative (reading and rereading text) self-regulated learning strategies (Thomas et al., 1993), and progress evaluation or self-monitoring (Butler & Winne, 1995; Kinzie, 1990; Pintrich & de Groot, 1990). Monitoring occurs after a self-regulated learner establishes a goal or set of goals against which progress is evaluated (Butler & Winne, 1995) and determines how to approach the learning task (Corno & Mandinach, 1983). This internal feedback loop appears to be unique to self-regulated learning. Because training and educational settings have goals or achievement requirements, it is logical that self-regulated students may perform better than engaged students. The question remains as to whether self-regulated students who are engaged perform best.

Another potential benefit of engagement, involvement, and self-regulated learning is

increased calibrated. People who are well calibrated are confident as well as accurate in their self-assessments of what they do and do not know. Zimmerman (1990) proposed that self-regulated learners are aware of whether they do or do not know something (i.e., are better calibrated), although the direct relation between engagement, involvement, or self-regulated learning and calibration has not been tested. Nevertheless, strategies for engagement, involvement, and/or self-regulated learning, such as increased processing (Maki, Foley, Kajer, Thompson, & Willert, 1990; Walczyk & Hall, 1989) and monitoring (Schraw, Potenza, & Nebelsick-Gullet, 1993) affect various levels of calibration, which suggests that strategies to improve calibration may exist as well as possible relationships between calibration and these three constructs.

Apparently, strategies that increase information processing tend to increase knowledge, which appears to enhance calibration (Maki et al., 1990; Walczyk & Hall, 1989). If increased knowledge makes tasks easier and calibration tends to occur because overconfidence is reduced with easy tasks (Bjorkman, 1992), it may be beneficial to entice students to increase their knowledge. Students given extra credit to improve calibration performed better and were less overconfident than students given extra credit to improve performance. Students who received incentives to improve calibration apparently shifted their attention from performance to monitoring, facilitating self-generated cognitive feedback (Schraw et al., 1993). Self-generated feedback was also suspected to develop when embedded questions and examples were used in texts (Walczyk & Hall, 1989). This suggests that feedback should include not only achievement, but calibration information. Hence, there is a need to determine the direct effects of engagement, involvement, and self-regulated learning on calibration.

Now that the constructs have been defined and their effects identified, correlates of these constructs will be discussed. Only self-interest (Wehlage, 1989) and affect (Skinner & Belmont, 1993) were found in the literature as correlates of engagement. Similar to engagement, few correlates of involvement have been identified. Academic involvement correlated with need for achievement (Farrell & Mudrack, 1992), task difficulty/goal (Reed & Schallert, 1993), Protestant work ethic (Farrell & Mudrack, 1992), importance of school activities (i.e., relevance; Farrell & Mudrack, 1992; Reed & Schallert, 1993), and positive affect (Reed & Schallert, 1993), but negligibly correlated with learned helplessness (Farrell & Mudrack, 1992).

The majority of literature focused on the correlates of self-regulated learning. From

numerous models and some correlational studies, the proposed correlates of self-regulated learning included confidence (Zimmerman, 1990), self-efficacy (Henderson, 1985; Kinzie, 1990; McMillan, Simonetta, & Singh, 1994; Pintrich & de Groot, 1990; Schunk, 1990; Zimmerman, 1990), self-esteem (McCombs & Marzano, 1990), self-concept (Howard-Rose & Winne, 1993; Thomas et al., 1993), learned helplessness (Henderson, 1986), locus of control (Henderson, 1986; Wilhite, 1990), strategy and capacity beliefs (Skinner, Wellborn, & Connell, 1990), earner or personal control (Kinzie, 1990), relevance (Kinzie, 1990), task importance (Pintrich & de Groot, 1990), curiosity (Kinzie, 1986) or interest (Pintrich & de Groot, 1990), motivation Kinzie, 1986; McCombs & Marzano, 1990), personal development goals (McCombs & Marzano, 1990), achievement goals (McCombs & Marzano, 1990; Schunk, 1990) or level of hallenge (Pintrich & de Groot, 1990), affect or mood (McCombs & Marzano, 1990), and prior chievement (Pintrich & de Groot, 1990). Correlates of engagement, involvement, and self-gulated learning are presented in Figure 2.

Categorizing similar correlates together reveals four groups of correlates relating to self-ncept, relevance, goals, and affect. Interestingly, motivation is conceived to be influenced by umber of correlates such as self-efficacy (Kinzie, 1990; McMillan, et al., 1994) personal ntrol, relevance, curiosity (Kinzie, 1990), self-esteem, personal development goals, affect, od (McCombs & Marzano, 1990), and attitudes (McMillan et al., 1994).

It appears as though engagement, involvement, and self-regulated learning all include the relates of relevance and affect. Yet, the majority of correlates of self-regulated learning tend eflect the likelihood that one will set challenging goals and monitor one's performance tive to those goals. That is, learners who have high perceptions of themselves (i.e., high ls of self-efficacy, confidence, self-esteem), identify relevance in the material (i.e., interest, osity), and experience positive affect are more likely to set challenging goals, and to monitor performance relative to those goals than learners with low self perceptions.

Given the potential positive effects of engagement, involvement, and self-regulated ing, it is necessary to identify good measures of these constructs. The most common form easurement is the questionnaire (e.g., Finn et al., 1991; Goff & Ackerman, 1992; Pintrich Groot, 1990; Reed & Schallert, 1993), which is used to collect self-reports (e.g., Skinner lmont, 1993; Skinner et al., 1990; Wilhite, 1990), or teacher ratings or observations (e.g., et al., 1991; Lee & Anderson, 1993). Usually students are asked to consider what they do

ıen studying, working on a task, or preparing for a test.

Current Measures

Because of space limitations, the current measures of engagement, involvement, and self-ulated learning are identified below, but thorough descriptions and the summary charts are :ussed in more detail in a related document (Stone & Bennett, 1996). Current measures of ıgement include Biggs' (1987) Learning Process Questionnaire (LPQ), to assess high school ents' beliefs about learning, and Study Process Questionnaire (SPQ), to assess the study :fs of college students. These questionnaires were designed to capture whether students have face, deep, or achieving approach to learning. These approaches, in turn, influence student ves and strategies, which are measured in the questionnaires.

Finn et al. (1991) created the Student Participation Questionnaire (SPQ). The 29-item was used by the classroom teacher to assess fourth grade students' level of participation, g concepts of engagement such as persistence, effort, attention, and interest. Finn et al.) identified the factors nonparticipatory behavior, effort, and initiative. Yet, these factors ot be as distinct as they could be given that the items had the highest loading on their ive factors, but also often loaded on other factors. Similarly, Skinner & Belmont (1993) udent and teacher reports to measure students' engagement levels. Third, fourth, and ade students and teachers assessed the students' behavioral (effort, attention, persistence) otional (interest, happiness, anxiety, anger) engagement.

Similarly, Thomas et al.'s (1993) Study Activity Questionnaire taps cognitive and effort nent study activities by assessing six areas: level of cognitive processing, tational level, initiative, autonomous management, memory augmentation, and effort ıent. Time spent studying was also assessed. It is important to note that there were onsistencies between the data reported in the text and tables. Nevertheless, the three hich tended to reflect engagement, level of processing, representational level, and were the only positively correlated dimensions with the achievement test for this study .30, respectively) (Thomas et al., 1993).

gagement has also been perceived as a personality trait (Goff & Ackerman, 1992). s construct on undergraduate students, this measure of intellectual engagement tapped :llectual engagement (e.g., engagement in thoughtful or intellectual pursuits) and introverted intellectual engagement (e.g., enjoying involved discussions;

pleasure in evaluating one's own thoughts and feelings), absorption (e.g., other people seem not to exist when concentrating), interest in arts and humanities, and interest is social science. Intellectual engagement was determined to be distinct from factors such as openness, directed activity (lack of distractibility and energy), science/technology interests, and conscientiousness (e.g., prefers had work) (Goff & Ackerman, 1992). This conscientiousness factor appears to reflect involvement which supports the proposition that engagement and involvement may be distinct constructs.

Effective time within an allotted time period during which students were actively participating in learning has been used to measure engagement (Kumar, 1991). If engaged students become so engrossed in their work that their perception of time becomes distorted (Goff & Ackerman, 1992), then time spent on task may be a crude measure of engagement. That is, students who spend more time on task may be more engaged in the material.

Unlike engagement measures, only one involvement scale was found. Reed and Schallert (1993) developed a questionnaire, which identified two dimensions of involvement, concentration and understanding. Concentration included items related to attention (high), task difficulty (moderate, but achievable), and importance (high) (Reed & Schallert, 1993). Involvement was found to be different than interest (Reed & Schallert, 1993). Yet, several of these items actually appear to be measures of engagement.

A review of involvement measures in other areas provided insight into appropriate involvement items. Goldsmith and Emmert (1991) found three consumer product involvement measures that measured personal qualities (inherent interests, values, needs), physical characteristics (characteristics that increase interest), situational conditions (temporarily increases relevance or interest toward object), perceived risk (importance), the rewarding nature of a product, and the ability of a brand to convey status, personality, and identity (Goldsmith & Emmert, 1991). These aspects may be classified into interest, relevance, and importance categories. Similarly, Farrell and Mudrack (1992) adapted two of the job and work involvement scales of Kanungo (1982) to measure academic involvement of older students.

Although several models of self-regulated learning have been proposed (e.g., Butler & Winne, 1995; Kinzie, 1990; McCombs & Marzano, 1990), no well developed measures of self-regulated learning have emerged. Usually, researchers develop scales for their specific studies, often tapping the fourteen self-regulated strategies identified by Zimmerman and Martinez-Pons

(1986). Interestingly, Corno, Collins, and Capper's (1982) Self-Regulated Learning Scale tapped five general areas of self-regulated learning strategies, deliberate alertness, selectivity, accessing schemata, planning, and monitoring, which also overlapped with the fourteen self-regulated learning strategies identified by Zimmerman and Martinez-Pons (1986).

Recently, Zimmerman, Bandura, and Martinez-Pons (1992) developed two scales for assessing learners' self-efficacy for self-regulated learning and for academic achievement. Self-efficacy appears to be an important aspect of self-regulated learning. Similarly, the more positively one self-assessed one's memory, the individual had a more positive self-concept and better achievement (Wilhite, 1990). The EMQ was also significantly related to two dimensions on Christopoulos, Rohwer, and Thomas's (1987) Study Activity Survey, uniform processing and the generation of interpreted information (Wilhite, 1990).

Finally, measures of time also appear to be important for self-regulated learning. Yet, instead of distorting their perception of time, self-regulated learners tend to be cognizant of and monitor time in order to complete all tasks.

Proposed Measures

The definitions, correlates, and current measures were combined into the following proposed measures of engagement, involvement, and self-regulated learning. Only descriptions of each measure follows, but example items are presented in Appendix A.

Engagement should tap six areas: interest, attention, absorption, persistence, effort, and level of cognitive processing. Engaged students should find topics to be inherently interesting, whereby their attention is directed to the subject matter, they are not easily distracted from the material, there is a distortion of time and a lack of awareness of their environment, and the students are absorbed in the material. An absorbed student would be persistent in the quest for nformation about the topic and would display effort in mastering the material, but not iecessarily relative to a specific goal.

The concept of involvement includes learning/school importance, commitment, onscientiousness, and responsibility/ownership. School/learning importance reflects the udents' perception of how relevant the topics are to them and whether schooling in general is nportant. Commitment is one's connection with the institution while conscientiousness includes ne's belief in hard work.

A self-regulated learner develops goals for the task, establishes an environment conducive

to learning, uses various processing techniques and/or strategies, and monitors the environment, strategies, and progress toward the goals. The self-regulated learning factors measured should include: goal development, planning, self-concept, strategies, and monitoring.

Once these measure have been tested to be reliable and valid, it is important to determine how to increase students' levels of engagement, involvement, and self-regulated learning, in order for them to spend the necessary time and effort in their studies to do well. Students need information which inspires them and helps them understand why the material they are learning is important, relevant, and meaningful (Guskey, 1988), which are correlates of engagement, involvement, and self-regulated learning. Yet, there is a difference between helping students find meaning and relevance in topics vs. entertainment (e.g., the Dr. Fox phenomenon or educational seduction), drama, or acting (Guskey, 1988).

It may be possible to increase engagement with various teaching or presentation styles (Guskey, 1988), or teacher interactions (Greenwood et al., 1994; Skinner & Belmont, 1993). Yet, due to the scarcity of research concerning instructor involvement with adult learners, it is unclear how strong an effect instructor involvement could have on increasing adult engagement or how much feedback a tutor might need to provide to increase student engagement.

Another, perhaps more realistic, means for increasing adult engagement, and calibration, is to increase students' depth of processing (Walczyk & Hall, 1989). Strategies to increase student learning can be taught to the students (Tobias, 1989); yet, the instructions must have a certain level of preciseness (Helstrup, 1989). Hence, another noninterpersonal means to promote engagement is with goals (Butler, 1993).

Because the majority of literature on involvement focused on teacher or parent involvement, there is little research on how to increase student involvement. Nevertheless, mentoring may be one means to affect student involvement (Jacobi, 1991).

Like engagement and involvement, self-regulated learning is also affected by teaching styles. Corno and Mandinach (1983) proposed that student cognitive processes are often squelched when students are given information and then told how to organize the information. In fact, teacher compensations (giving sample items, telling students what to study, reducing or eliminating study demands, extra credit) tended to be negatively related to achievement (Thomas et al., 1993). Therefore, it is important to develop instructional materials which support but push the student to next level of comprehension (Henderson, 1986). That is, tutorial assistance

is like a scaffolding which is related to the students' level of comprehension, which should help transfer the responsibility of learning to the learner (Henderson, 1986). Even if the student is given a scaffolding from which to work, self-regulated learning may be a threat to underachievers' self-concept (Paris & Newman, 1990). Thus, interactions with teachers or instructional feedback may also affect students' levels of self-regulated learning (Skinner et al., 1990).

Teacher interaction may also affect monitoring, a central strategy to self-regulated learning (Butler & Winne, 1995). Students often are not good at self-monitoring (Helstrup, 1989; Kinzie, 1990). Hence, students need some type of guidance perhaps mentoring, advisement, and/or training (Kinzie, 1990). Feedback, though, should also reduce this inaccuracy. Feedback about what strategies could be used to increase learning and what strategies were or were not used, probably helps students recognize cognitive activities they perform while learning, and in turn enhances calibration (Butler & Winne, 1995).

Similarly, it may also be possible to increase self-monitoring by informing students as to what calibration is and motivating them to be better calibrated (Schraw et al., 1993). Perhaps giving feedback, getting students to focus on calibration, and embedding questions and examples in the text help students set realistic goals. If task difficulty or goals vary, this may explain why students' levels of self-regulation change for various tasks (Howard-Rose & Winne, 1993). As Pintrich and de Groot (1990) found, involvement or self-regulation changed as students progressed through various phases of writing a paper. If goals can be set whereby they push students to their next level of comprehension, there is a greater chance that the responsibility of learning will be transferred to the student (Henderson, 1986).

Proposed Research

All three constructs are proposed to be distinct from each other, although certain facets of the constructs may overlap. Ideally, the best situation exists when one is a self-regulated learner and engaged. Involvement is expected to have less effect. Therefore, to begin, it is necessary to develop a reliable and validated measure of engagement, involvement, and self-regulated learning. Given that much of the research has focused on elementary or secondary education populations, the constructs and measures of engagement, involvement, and self-regulated learning need to be evaluated relative to an adult population. These measures also need to be assessed in both educational and non-educational settings to examine generalizability.

To test these proposed measures, the scales would be administered to adult populations at the beginning and end of an academic and/or training term. Factor analysis would be used to identify the factor structures of these measures, to determine if the factor structures are similar across time, and to assess whether individuals' levels of these constructs change over time. Final grades would also be collected. These scores would be used to determine the relationship between the measures and achievement.

Once these constructs, or something similar, can be reliably measured, how the constructs relate to measures of calibration will be evaluated. Confidence measures before tests (after studying) will be collected. It may be too invasive to collect confidence ratings during tests. It may also be beneficial to collect confidence ratings for each test during the term to evaluate the stability of the measures. Again, final grades will be needed to calculate calibration.

Next, because a number of the influences of engagement, involvement, and self-regulated learning were taken from proposed models of self-regulated learning or were tested relative to the child population, empirical tests need to be conducted to determine what teaching techniques or strategies may actually influence engagement, involvement, and/or self-regulated learning. Specifically, the effects of feedback (about strategy use, calibration) and goals, including the notion of scaffolding, will be investigated. Additionally, it may be possible to increase self-generated feedback with imbedded exercises such as questions and examples. The types of imbedded exercises that are helpful may depend on the competence level of the student. Therefore, the scaffolding approach for presenting exercises to the students and goal setting would again be investigated. This information could be useful in restructuring stand up lecture and in the development of computerized classroom presentations.

Instructor involvement with the students is also related to mentoring. Although elementary aged school children were influenced by the attention their teachers gave them, no studies have evaluated these effects on adults. If teacher involvement is important to adult learners, then more interaction may need to be instituted into computerized classrooms.

It is also necessary to determine how students develop ownership of their learning. Theoretically, self-regulated learners accept greater responsibility for their achievement outcomes (Zimmerman, 1990). Also, learners who have had the opportunity to assess their own academic work have gained more from their learning experience (Dickinson, 1992; Sereda, 1993).

Students, who were teachers enrolled in a course which offered collaborative assessment, became more autonomous over time and became better assessors of their own work (Dickinson, 1992). Thus, when strict performance criteria are set, and students are allowed to collaboratively assess their own work, the students tend to become more invested in their learning. More research is needed to determine how self-assessment or ownership may be enhanced and how self-assessment and ownership affect learning.

Finally, although a small number of studies were longitudinal and evaluated engagement, involvement, and/or self-regulated learning relative to achievement tests, the majority of research focused on course tests, quizzes, or end of the course achievement tests. Thus, the long term effects of engagement, involvement, and self-regulated learning need to be evaluated.

References

Ackerman, P. L. & Goff, M. (1994). Typical intellectual engagement and personality: Reply to Rocklin (1994). Journal of Educational Psychology, 86, 150-153.

Ainley, M. D. (1993). Styles of engagement with learning: Multidimensional assessment of their relationship with strategy use and school achievement. Journal of Educational Psychology, 85, 395-405.

Biggs, J. B. (1987). Student approaches to learning and studying. Melbourne, Australia: Australian Council for Educational Research.

Bjorkman, M. (1992). Knowledge, calibration, and resolution: A linear model. Organizational Behavior and Human Decision Processes, 51, 1-21.

Butler, R. (1993). Effects of task- and ego-achievement goals on information seeking during task engagement. Journal of Personality and Social Psychology, 65, 18-31.

Butler, D. L. & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. Review of Educational Research, 65, 245-281.

Christopoulos, J. P., Rohwer, W. D. Jr., & Thomas, J. W. (1987). Grade level differences in students' study activities as a function of course characteristics. Contemporary Educational Psychology, 12, 303-323.

Corno, L., Collins, K., & Capper, J. (1982). Where there's a way there's a will: Self-regulating the low-achieving student. ERIC Document ED 222 499.

Corno, L. & Mandinach, E. B. (1983). The role of cognitive engagement in classroom learning and motivation. Educational Psychologist, 18, 88-108.

Dickinson, L. (1992). Collaboration in assessment: Empowering the individual course
nber. In E. Sadtono (Ed.) Language teacher education in a fast-changing world. Anthology
:s 29. Eric Document ED 369 275.

Farrell, G. M. & Mudrack, P. E. (1992). Academic involvement and the nontraditional
:nt. Psychological Reports, 71, 707-713.

Finn, J. D., Folger, J., & Cox, D. (1991). Measuring participation among elementary
: students. Educational and Psychological Measurement, 51, 393-402.

Goff, M. & Ackerman, P. L. (1992). Personality-intelligence relations: Assessment of
ıl intellectual engagement. Journal of Educational Psychology, 84, 537-552.

Goldsmith, R. E. & Emmert, J. (1991). Measuring product category involvement: A
rait-multimethod study. Journal of Business Research, 23, 363-371.

Greenwood, C. R., Terry, B., Marquis, J., & Walker, D. (1994). Confirming a
mance-based instructional model. School Psychology Review, 23, 652-668.

Guskey, T. R. (1988). Improving student learning in college classrooms. Springfield, IL:
s C. Thomas, Publisher.

Helstrup, T. (1989). Active and passive memory: States, attitudes, and strategies.
ıavian Journal of Psychology, 30, 113-133

Henderson, R. W. (1986). Self-regulated learning: Implications for the design of
ional media. Contemporary Educational Psychology, 11, 405-427.

Ioward-Rose, D. & Winne, P. H. (1993). Measuring component and sets of cognitive
:s in self-regulated learning. Journal of Educational Psychology, 85, 591-604.

acobi, M. (1991). Mentoring and undergraduate academic success: A literature review.
of Educational Research, 61, 505-532.

anungo, R. N. (1982). Measurement of job and work involvement. Journal of Applied
gy, 67, 341-349.

inzie, M. B. (1990). Requirements and benefits of effective interactive instruction:
ontrol, self-regulation, and continuing motivation. Educational Technology Research
lopment, 38, 5-21.

ımar, D. D. (1991). A meta-analysis of the relationship between science instruction
ıt engagement. Educational Review, 43, 49-61.

Lee, O. & Anderson, C. W. (1993). Task engagement and conceptual change in middle school science classrooms. American Educational Research Journal, 30, 585-610.

Maki, R. H., Foley, J. M., Kajer, W. K., Thompson, R. C., & Willert, M. G. (1990). Increased processing enhances calibration of comprehension. Journal of Experimental Psychology: Learning, Memory, and Cognition, 16, 609-616.

McCombs, B. L. & Marzano, R. J. (1990). Putting the self in self-regulated learning: The self as agent in integrating will and skill. Educational Psychologist, 25, 51-69.

McCombs, B. L. & Whisler, J. S. (1989). The role of affective variables in autonomous learning. Educational Psychologist, 24, 277-306.

McMillan, J. H., Simonetta, L. G., & Singh, J. (1994). Student opinion survey: Development of measures of student motivation. Educational and Psychological Measurement, 54, 496-505.

Paris, S. G. & Newman, R. S. (1990). Developmental aspects of self-regulated learning. Educational Psychologist, 25, 87-102.

Pintrich, P. R. & de Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. Journal of Educational Psychology, 82, 33-40.

Reed, J. H. & Schallert, D. L. (1993). The nature of involvement in academic discourse tasks. Journal of Educational Psychology, 85, 253-266.

Schraw, G., Potenza, M. T., & Nebelsick-Gullet, L. (1993). Constraints on the calibration of performance. Contemporary Educational Psychology, 18, 455-463.

Schunk, D. H. (1990). Goal setting and self-efficacy during self-regulated learning. Educational Psychologist, 25, 71-86.

Sereda, J. (1993). Educational quality indicators in art and mathematics. Alberta Journal of Educational Research, 39, 217-233.

Skinner, E. A. & Belmont, M. J. (1993). Motivation in the classroom: Reciprocal effects of teacher behavior and student engagement across the school year. Journal of Educational Psychology, 85, 571-581.

Skinner, E. A., Wellborn, J. G., & Connell, J. P. (1990). What it takes to do well in school and whether I've got it: A process model of perceived control and children's engagement and achievement in school. Journal of Educational Psychology, 82, 22-32.

Stone, N. J. & Bennett, W. R., Jr. (1996). Conceptual development of engagement, involvement, and self-regulated learning measures (Report No. AL/HR-TP-1996-DRAFT).

Thomas, J. W., Bol, L., Warkentin, R. W., Wilson, M., Strage, A., & Rohwer, W. D. Jr. (1993). Interrelationships among students' study activities, self-concept of academic ability, and achievement as a function of characteristics of high-school biology courses. Applied Cognitive Psychology, 7, 499-532.

Tobias, S. (1989). Another look at research on the adaptation of instruction to student characteristics. Educational Psychologist, 24, 213-227.

Walczyk, J. J. & Hall, V. C. (1989). Effects of examples and embedded questions on the accuracy of comprehension self-assessments. Journal of Educational Psychology, 81, 435-437.

Wehlage, G. G. (1989). Engagement, not remediation or higher standards. In J. M. Lakebrink (Ed.) Children at Risk. Springfield, IL: Charles C. Thomas, Publisher.

Whitener, E. M. (1989). A meta-analytic review of the effect on learning of the interaction between prior achievement and instructional support. Review of Educational Research, 59, 65-86.

Wilhite, S. C. (1990). Self-efficacy, locus of control, self-assessment of memory ability, and study activities as predictors of college course achievement. Journal of Educational Psychology, 82, 696-700.

Zimmerman, B. J. (1986). Becoming a self-regulated learner: Which are the key subprocesses? Contemporary Educational Psychology, 11, 307-313.

Zimmerman, B. J. (1990). Self-regulated learning and academic achievement: An overview. Educational psychology, 25, 3-17.

Zimmerman, B. J., Bandura, A., & Martinez-Pons, M. (1992). Self-motivation for academic attainment: The role of self-efficacy beliefs and personal goal setting. American Educational Research Journal, 29, 663-676.

Zimmerman, B. J. & Martinez-Pons, M. (1986). Development of a structured interview for assessing student use of self-regulated learning strategies. American Educational Research Journal, 23, 614-628.

Figure 1:

Descriptors of Engagement, Involvement, and Self-Regulated Learning

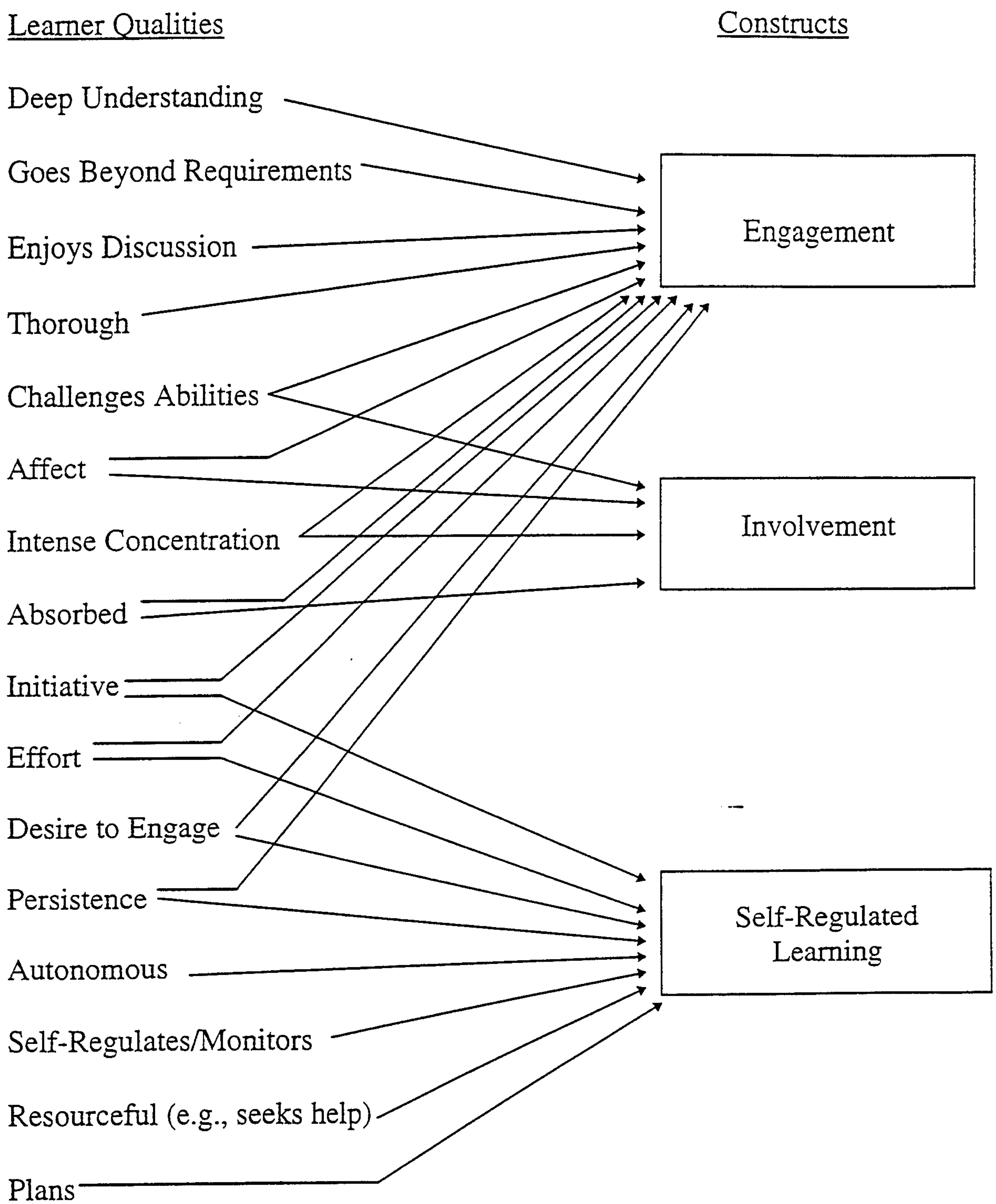


Figure 2:

Correlates of Engagement, Involvement, and Self-Regulated Learning

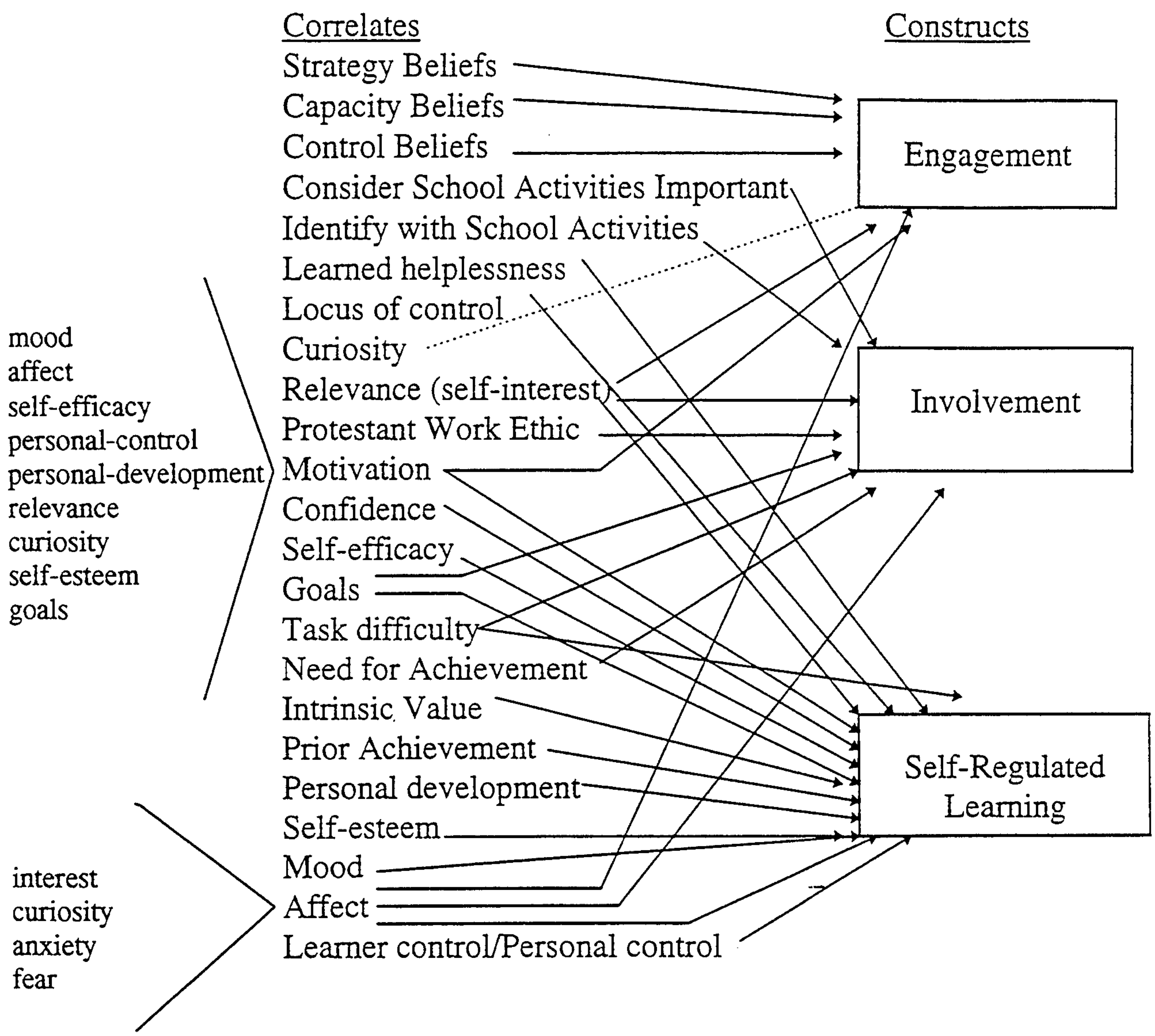


Table 1: Examples of Proposed Measurement Items

Use the following scale to rate the extent to which the following are true about yourself: (1=never, 2=almost never, 3=rarely, 4=sometimes, 5=usually, 6=almost always, 7=always)

Proposed Engagement Items

Attention
1. It is easy to focus my attention to the subject material.
2. When working on the material my minds seems to wander.
3. When working on the material I have intense concentration

Absorption
1. I tend to lose track of time.
2. I become completely absorbed in what I am doing.
3. I do not keep track of the time spent working on the material, I just take as much time as needed to finish.

Interest
1. I seek more information related to interesting topics, even if it takes me off task.
2. I participate actively in discussions.
3. Any topic can be highly interesting once I get into it.

Effort
1. I enjoy having problems or puzzles in life to solve.
2. It is fun to engage in problem solving.
3. I prefer to work easy rather than hard problems.

Persistence
1. I try to finish assignments even when they are difficult
2. When problems are difficult, I easily get discouraged and stop trying.
3. I work beyond the course requirements to learn the material thoroughly.

Level of Cognitive Processing
1. I easily can distinguish the important from the unimportant information?
2. I integrate new information with what I already know.
3. I try to apply my knowledge to other situations.

Proposed Involvement Items

Learning/School Importance
1. It is important to choose courses that will get me a good job, not because they are interesting.
2. School is important.

Commitment
1. I shouldn't be expected to spend time studying what everyone knows will not be on the test.
2. I am at school mainly because I will get a better job with more education.
3. I have a strong work ethic toward my studies.

Conscientiousness
1. I enjoy hard work.
2. I am willing to devote all my attention to my studies

Responsibility/Ownership
1. I take complete responsibility for own education.
2. I should be involved in my assessment/grading.

Proposed Self-Regulated Learning Items

Goal Development
1. I set goals for each study session.
2. It is important to establish goals before beginning to study.

Planning
1. I arrange a place to study that will be without distractions.
2. I determine, before studying, how to deal with possible interruptions.
3. I determine ahead of time how long I will work on a particular topic.

Self-Concept
1. I am confident in my ability to do well.
2. I have skills for overcoming course difficulties.
3. I can succeed in any course.

Strategies
1. I choose courses in which I know I can get a good grade.
2. I seek the necessary information for doing well in the course.
3. I review my notes, tests, and textbooks.

Monitoring
1. I monitor the amount of time I spend on a topic.
2. I continuously evaluate whether I will meet my set goals for the course.
3. After studying, I test myself to determine if I have successfully mastered the material.

Feedback
1. I use test results and my own evaluations to determine what I need to do differently to succeed in the course.

APTITUDE-ATTRIBUTE INTERACTIONS IN TEST PERFORMANCE

Brenda Sugrue
Assistant Professor
College of Education

University of Iowa
304N Lindquist Center
Iowa City, IA 52241

Final Report for:
Summer Faculty Research Program
Armstrong Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington DC

and

Armstrong Laboratory

October 1996

APTITUDE-ATTRIBUTE INTERACTIONS IN TEST PERFORMANCE

Brenda Sugrue
Assistant Professor
College of Education
University of Iowa

Abstract

This study examined the extent to which test performance is jointly influenced by attributes of test items and general aptitude of test takers. A methodology was developed for coding item attributes and simultaneously analyzing aptitude-attribute, attribute-attribute, and main effects. Pretest and posttest data from a previous study were reanalyzed using this methodology. A number of aptitude-treatment interactions and attribute-attribute interactions were found. For example, on the posttest, the lower one's level of general ability, the lower one's performance on items requiring generation of responses, compared to performance on items requiring selection of responses. Regardless of aptitude, items requiring symbolic knowledge were easier in selection format than in generation format, whereas items requiring procedural skill were (at least on the posttest) equally easy regardless of format. The existence of such interactions means that conclusions about item difficulty based only on main effects may be misleading.

APTITUDE-ATTRIBUTE INTERACTIONS IN TEST PERFORMANCE

Brenda Sugrue

Introduction

It is widely acknowledged that test item difficulty and observed variability in performance are a function of the complex interaction of a host of test and test-taker characteristics (Hunt, 1995; Messick, 1995; Nichols, 1995; Snow & Lohman, 1989; Swanson, Norman, & Linn, 1995). However, most studies of the influence of item characteristics on difficulty have examined only main effects (e.g., Allen et al., 1995; Tatsuoka, 1995; Martines & Katz, 1992). Consequently, we know little about the extent to which the effects of item attributes depend on individual differences of students, or the extent to which attributes interact with each other in their effects on performance. For example, some item formats may be more difficult for some students than others, or some formats may be more difficult for all students when testing one particular type of knowledge. These kinds of interaction effects were the focus of this study.

One of the goals of the study was to develop a generalizable methodology for investigating aptitude-attribute interactions. The methodology involves the assignment of domain-independent attribute codes to items, and the use of repeated measures analysis of variance to examine all possible interaction and main effects simultaneously. The methodology can be applied to any dataset that has domain-specific test item data and aptitude data for individual students. This paper describes the application of the methodology to one dataset which had data on the same test items administered as a pretest and posttest. Analysis of both pretest and posttest data permitted examination of the extent to which interaction effects might hold regardless of amount of domain-specific knowledge. The domain for this study was descriptive statistics. Instruction was provided by the Stat Lady intelligent tutor developed by Shute & Gluck (1994). Subjects were all trainees in the United States Air Force.

Item attributes

Many aspects of test items can vary. Item stimulus characteristics such as amount of verbal content, amount of scaffolding, amount of irrelevant information, and medium of presentation can vary. Aspects of the response elicited by the item can vary; for example, the format, authenticity, length and complexity of the response can vary. Other attributes of items that can vary and that contribute to their difficulty are scoring methods (Baxter, Glaser, & Raghaven, 1993) and similarity to instruction (Moody, 1996).

A variety of lists of item attributes have been generated or applied in previous studies. These lists are usually test- and domain-specific and do not distinguish among stimulus, response, scoring, and similarity to instruction characteristics. For example, Yepes-Baraya and Allen (in press) identified thirty-six attributes of science items on the National Assessment of Educational Progress (NAEP). These attributes accounted for 64% of variance in scores. Based on exploratory factor analysis, the attributes were clustered into five categories: content knowledge, reasoning and explaining, hypothesis formulation and testing, processing figural information, and item format and reading difficulty.

In a study of the cognitive processes elicited by mathematics and science tests used in the National Educational Longitudinal Study (NELS), Hamilton, Nussbaum, and Snow (1995) identified six categories of cognitive demands and generated questions that could be used to evaluate items on each category. Those categories were demands on working memory, use of language and communication, metacognitive skill demands, application of prior knowledge and expectations, acquisition of new knowledge, and use of scientific processes.

Attributes can be rated as being present or absent in an item, or as being present to varying degrees. Most studies, which have examined the extent to which attributes predict difficulty, have used dichotomous coding. Some attributes are easy to rate (e.g., response format, or scoring procedures). Other attributes are more difficult to rate; for example, identification of the cognitive processing demands of items requires examination of student work and/or observation and interview as students work on the items. Similarity to instruction attributes require review of instructional materials and interviews with teachers.

For this study, a small number of response attributes were selected to meet the following criteria:

1. the attribute could apply across a variety of tests and domains,
2. the attribute could be identified reliably from examination of the item stimulus, not actual response protocols, and
3. the attribute could be coded reliably by raters who are not experts in the domain being tested.

If aptitude-attribute interactions were found for this set of attributes, then it would legitimize aptitude-attribute interaction as a fruitful avenue for investigating sources of variability in test performance, much like early aptitude-treatment interaction studies legitimized that paradigm for research on instructional variables. Future studies could include a greater range of item attributes.

The specific attributes selected for this study were response format, type of knowledge required to respond, and type of cognitive processing required to respond. Response format could be selection or generation depending on whether students had to select a response from a set of options or generate a response. Type of knowledge required could be symbolic knowledge (SK), procedural skill (PS), or conceptual knowledge (CK), based the three-way distinction made by Shute (1995). Items requiring symbolic knowledge ask students to select or generate factual information about terms, symbols, rules and definitions. Items requiring procedural skill ask students to perform some sequence of actions and/or decisions. Items requiring conceptual knowledge ask students to predict or explain outcomes in terms of principles governing the causal relationships among concepts.

Shute (1995), in an evaluation study of the Stat Lady intelligent tutor , found that the gains made on different types of items from pretest to posttest were different for high and low aptitude students. High ability students made similar gains across all item types, while low aptitude students made dramatic gains on items measuring procedural skills, but very little gain on items measuring symbolic knowledge. Shute attributed this finding to the fact that students had more opportunity in the Stat Lady tutor to practice procedural skills and, because of a ceiling effect for the high aptitude students, low students appeared to gain most.

Finally, two types of cognitive processing were distinguished: retrieval and reasoning. An item was coded as requiring retrieval if it merely called on the student to retrieve some information that already exists in long-term memory. When an item was coded as requiring reasoning, then it was assumed that all test takers would engage in reasoning when responding. The exception would be a student who has practiced the exact same item extensively in the past. Items were coded according to the highest level of knowledge and cognitive processing required.

<u>Aptitudes</u>

A variety of individual differences could potentially interact with item characteristics to produce different score profiles. For example, student perceptions of difficulty, preferences for item types, as well as domain-general and domain-specific abilities could influence how a particular student interprets and responds to an item. For this study, only students' general ability was included as an aptitude variable. If such a broad aptitude variable were found to interact with item attributes, then it would justify proceeding to examine more differentiated aptitude variables in future studies.

Methodology

Existing dataset and additional variables

The dataset used in this study contained performance data for 104 subjects on 65 items measuring knowledge of descriptive statistics before and after instruction via the Stat Lady intelligent tutor (Shute, 1995). The dataset also contained a variable representing general cognitive ability; this variable was a factor score based scores on a number of tests from the computerized Cognitive Abilities Measurement (CAM-4) battery (Kyllonen et al., 1990). Additional variables, which represented mean scores on subsets of items manifesting particular attributes and combinations of attributes, were added to each dataset. Subsets of items were identified by coding the format (selection or generation), highest type of knowledge required (SK, PS, or CK), and highest cognitive processing demand (retrieval or reasoning), for each item. This coding was done independently by two raters, with the few discrepancies (less than 5%) in ratings being resolved through discussion.

Based on this coding, no items on the test were deemed to require conceptual knowledge. Thus, each of the three attribute variables in this study had two levels, which would have conveniently rendering the repeated measures design a 2 X 2 X 2, had there not been three empty cells. Table 1 shows the distribution of items across the three attribute variables (knowledge type, processing type, and format). The empty cells in the symbolic knowledge row indicate that, in this test, there were no items measuring reasoning about symbolic knowledge (in either format). Given the definition of symbolic knowledge (factual information), it is difficult to imagine an item that would require reasoning to generate or select some symbolic knowledge. The third empty cell in this test indicates that there were no items measuring retrieval of procedural skill in selection format. However, one could imagine situations where an item would require retrieval of procedural skill in a selection format; for example, if the item called for selection of the most appropriate sequence of actions to accomplish some goal, or the selection of the most appropriate procedure from a number of procedures.

Table 1. Number of items in each combination of attributes coded.

	Processing Type			
	Retrieval		Reasoning	
Knowledge Type	Selectio n Format	Generatio n Format	Selection Format	Generation Format
Symbolic Knowledge	32	6	0	0
Procedural Skill	0	5	12	10

Because this study was conducted on a set of items that were not consciously created to span the range of attributes under investigation, a number of compromises had to be made. Ideally, one should have equal numbers of items in every cell of a design representing all combinations of attributes. Because of the empty cells, and unequal n's in other cells, the processing attributes were dropped. This reduced the design to a 2 X 2 design, the two attributes being knowledge ant format. Even then, the problem of disproportionality (unequal numbers of items in each cell) remained, as shown in Table 2. To reduce the impact of this disproportionality on tests of statistical significance, the proportion of correct responses to items in each cell were computed rather than mean scores on sets of items in cells. The proportion correct variables were used in all analyses.

Table 2. Number of items in each combination of attributes coded for format by knowledge design.

	Format	
Knowledge	Selection	Generation
Symbolic	32	6
Procedural	12	15

<u>Analysis</u>

Repeated measures analysis of variance with a covariate was run using the General Linear Model procedure available in SPSS for Windows, Release 7.0. The General Linear Model procedure permits the simultaneous estimation of aptitude-attribute interactions, attribute-attribute interations, and main effects. Separate analyses were run for pretest and posttest data. The full model for the knowledge by format design included the repeated measures attribute variables knowledge and format as main effects, the two-way interaction between these, and the interaction of the "covariate" (i.e., general ability) with each of these main and interaction effects.

Whenever an aptitude-attribute interaction was found, the regression lines were plotted to show how scores were differentially related to aptitude, just as one would do in an aptitude-treatment interaction study (Cronbach & Snow, 1977). If aptitude did not interact with attributes, but attributes interacted with each other, then significantly different mean scores on subsets of items were plotted. If there were no interactions at all (either with or without the covariate), but a main effect existed, then the mean scores for the sets of items that represented the main effect were plotted. Had the design contained three attribute variables, the full model would have consisted of three main effects, all two-way interactions among them, the three-way interaction among the attributes, and the interaction of the aptitude variable with all combinations of attributes.

Results

Analysis of Variance Summary

Tables 3 and 4 display the pretest and posttest analysis of variance results for the model that included knowledge and format as repeated measures variables, and general ability as the continuous aptitude variable. The significance levels for the F-tests of effects indicate a number of aptitude-attribute interactions and attribute-attribute interactions. On the pretest, general ability interacted with knowledge, $F(1, 102)=14.8$, $p<.000$, and with format $F(1, 102)=6.56$, $p=.012$, but on the posttest general ability only interacted with format $F(1, 102)=4.53$, $p=.036$.

On both pretest and posttest, regardless of general ability, knowledge and format interacted with each other in their effects on performance, pretest $F(1,102)=20.369$, $p<.000$, posttest $F(1,102)=19.48$, $p<.000$. There was a main effect for format on both pretest and posttest, and a main effect for knowledge type on the posttest only. However, given the interaction effects found, these main effects must be qualified. The main effect of general ability was significant on both pretest performance, $F(1, 102)=53.34$, $p<.000$, and posttest performance, $F(1, 102)=91.87$, $p<.000$. The significant interaction effects will now be interpreted in detail.

Table 3. Pretest: Tests of within-subjects effects.

Source	Sum of Squares	df	Mean Square	F	Sig
K * F	.287	1	.287	20.369	.000
K * F * G	3.788E-02	1	3.788E-02	2.685	.104
Error(K *F)	1.439	102	1.411E-02		
K	5.058E-02	1	5.058E-02	2.158	.145
K * G	.347	1	.347	14.804	.000
Error(K)	2.390	102	2.344E-02		
F	5.097	1	5.097	270.784	.000
F * G	.124	1	.124	6.564	.012
Error(F)	1.920	102	1.882E-02		

Note: K=Knowledge type; F=Format; G=General ability

Table 4. Posttest: Tests of within-subjects effects.

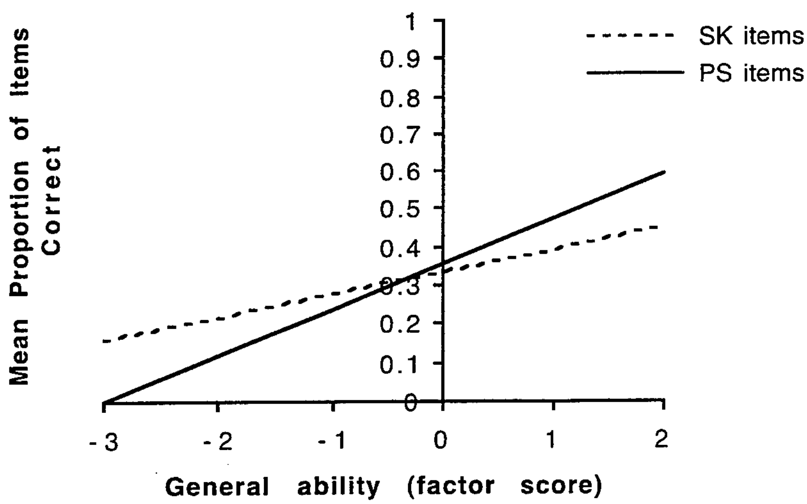
Source	Sum of Squares	df	Mean Square	F	Sig.
K * F	.301	1	.301	19.479	.000
K * F * G	3.158E-02	1	3.158E-02	2.045	.156
Error(K * F)	1.575	102	1.544E-02		
K	1.258	1	1.258	52.854	.000
K * G	1.521E-02	1	1.521E-02	.639	.426
Error(K)	2.428	102	2.381E-02		
F	.735	1	.735	28.976	.000
F* G	.115	1	.115	4.533	.036
Error(F)	2.587	102	2.537E-02		

Note: K=Knowledge type; F=Format; G=General ability

Aptitude-Treatment Interactions

On the pretest, aptitude was more strongly related to performance (proportion of items correct) on items requiring procedural knowledge (r=.58) than items requiring symbolic knowledge (r=.43). Figure 1 shows the regression lines for SK and PS items based on the equations Y=.336+.059X for SK and Y=.358+.119X for PS. The aptitude factor scores ranged from -2.92 to 1.86 with a mean of .06 and a standard deviation of .96. The interaction is disordinal; at the lowest end of the general ability scale, performance was better on SK items than on PS items, but at the high end of the scale, students did better on PS items than SK items. There was no difference between the mean scores on SK and PS items, as indicated by the closeness of the intercepts in Figure 1.

Figure 1. Pretest: Regression lines showing knowledge-type by aptitude interaction.



On the pretest, aptitude was more strongly related to performance on items requiring selection of responses (r=.62) than items requiring generation of responses (r=.43). Figure 2 shows the regression lines for selection and generation items based on the equations Y=.458+.107X for selection and Y=.236+.071X for generation. The interaction is ordinal; for all levels of ability, performance on selection items was better than on generation items. Indeed, the main effect of format was significant, with the mean proportion of items correct being .46 for selection items and only .24 for generation items. However, the gap between performance on items of different formats is wider at the higher end of the general ability scale than at the lower end.

On the posttest, aptitude was also more strongly related to performance on selection items (r=.69) than generation items (r=.62). Figure 3 shows the regression lines for selection and generation items on the posttest based on the equations Y=.613+.131X for selection and Y=.529+.166X for generation. As with the aptitude-format interaction on the

pretest, this interaction is ordinal, but in a different way. Again, the mean proportions of correct answers were significantly different: .61 for selection items and .53 for generation items. However, in this case, the higher the general ability of the student, the more similar was performance on the two types of items.

Figure 2. Pretest: Regression lines showing format by aptitude interaction.

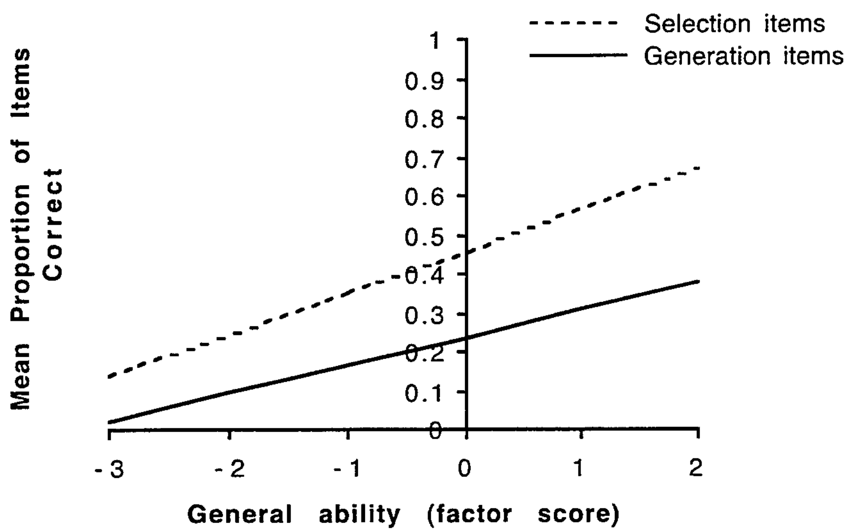
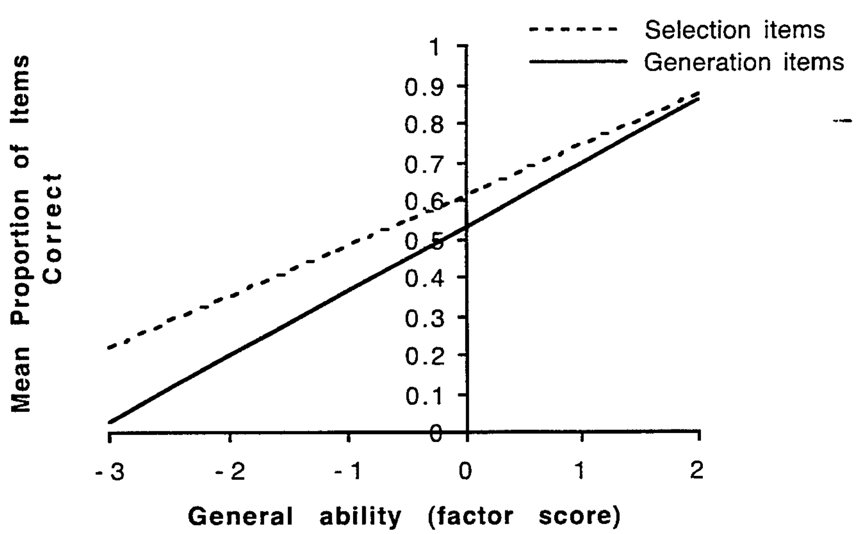


Figure 3. Posttest: Regression lines showing format by aptitude interaction.



Attribute-Attribute Interactions

On both the pretest and posttest, performance on items requiring different types of knowledge depended on the format of the items. While in both tests, performance on selection items was higher than performance on generation items (the main effect of format was significant on both tests), this discrepancy was greatest for items requiring symbolic knowledge, particularly on the pretest. Figures 4 and 5 show this format by knowledge type interaction. On the posttest, performance on PS items was higher than performance on SK items (this main effect was significant); however, this gap was only significant for items in generation format.

Figure 4. Pretest: Mean scores depicting knowledge-type by format interaction.

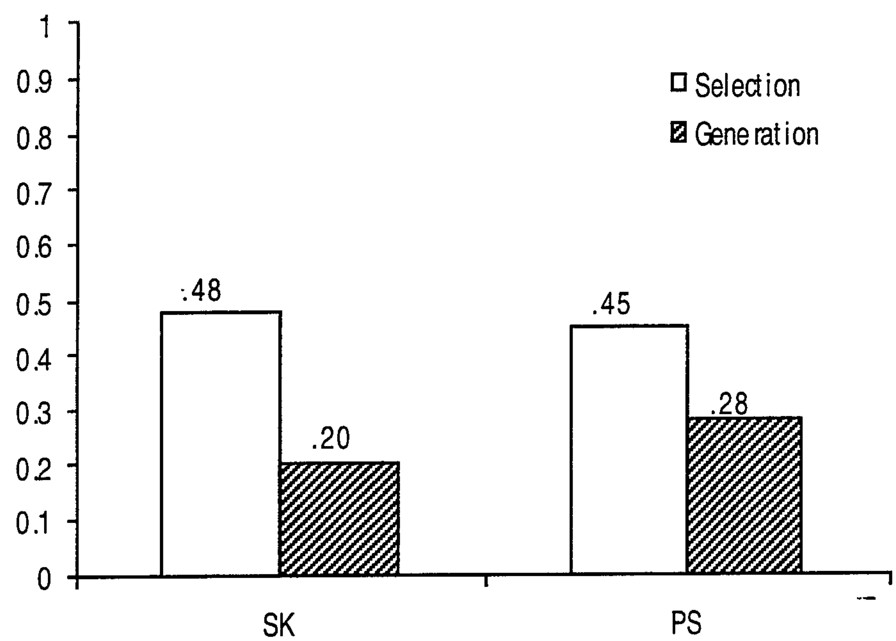
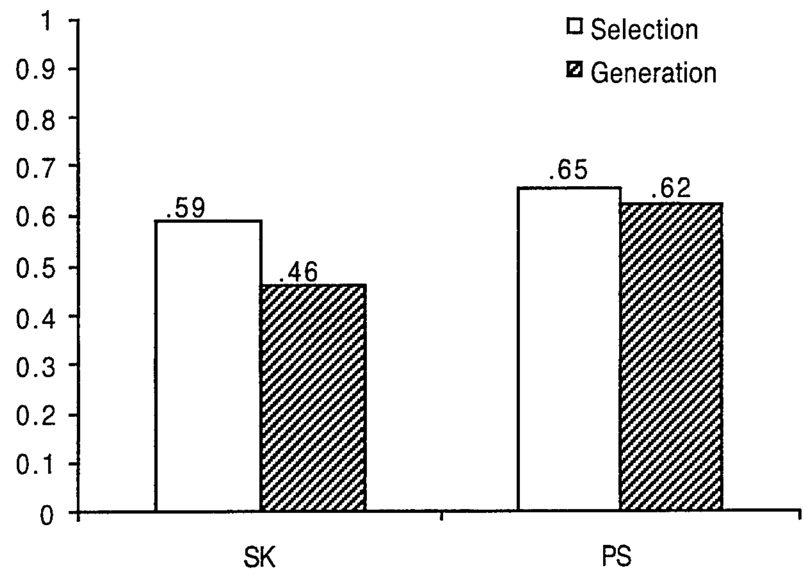


Figure 5. Posttest: Mean scores depicting knowledge-type by format interaction.



Discussion

If this study had not adopted an interactionist perspective, the conclusions, based only on an examination of main effects, would have been:

1. items measuring symbolic knowledge and procedural knowledge were equally difficult on the pretest, but on the posttest items measuring procedural skill were easier than items measuring symbolic knowledge;
2. selection items were easier than generation items on both pretest and posttest.

These conclusions would have masked important differences in the relative difficulty of items for different students, and important interactions between format and type of knowledge being measured. The conclusions based on this study contradict the "main effects" conclusions in the following ways:

1. items measuring symbolic knowledge and procedural knowledge on the pretest were NOT equally difficult for all students. Students with low general ability found PS items more difficult than SK, and students with higher general ability found SK items more difficult than PS item;
2. PS selection items were NOT easier than SK selection items on the posttest;
3. selection items were NOT easier than generation items on the posttest for higher ability students;
4. selection items measuring PS were NOT easier than generation items measuring PS on the posttest.

The fact that significant interactions were found with the gross measure of aptitude and the small number of item attributes used in this study indicates that this approach to investigating sources of variance in test performance is worth pursuing. Further studies need to be conducted to determine if any of the interactions reported here were artifacts of the particular set of test items studied, or the particular instructional emphasis in the Stat Lady learning environment. If consistent aptitude-attribute interactions and attribute-attribute interactions were to be found across tests in different domains, or across students with different amounts and types of instructional experiences in a domain, then it would be possible to predict the relative difficulty of any test item for any student. One could use information about students' general abilities and information about their instructional exposure to create "macroadaptive" tests that would more accurately and efficiently generate estimates of students' knowledge of a domain.

Meanwhile, just knowing that aptitude-attribute interactions and attribute-attribute interactions exist justifies the development of tests that contain a wide variety of types of items, so that students are given maximum opportunity to demonstrate whatever knowledge they possess. One would also be justified in generating more fine-grained reports of test performance; knowing that a student did well on items of one type but not on items of another type is more useful diagnostic and predictive information than having a single-score estimate of the student's knowledge.

References

Allen, N.L., Park, C., Liang, J., & Thayer, D. (1995). *Relationships between test specifications, item responses, task demands and item attributes in large-scale assessment.* Paper presented at the 1995 Annual Meeting of the American Educational Research Association, San Francisco.

Baxter, G. P., Glaser, R., & Raghavan, K. (1993). *Analysis of cognitive demand in selected alternative science assessments.* (CSE Tech. Report). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Cronbach, L.J., & Snow, R.E. (1977). *Aptitudes and instructional methods.* New York: Irvington.

Hamilton, L.S., Nussbaum E.M., & Snow, R.E.(1995). *Alternative interview procedures for validating science assessments.* Paper presented at the 1995 Annual Meeting of the American Educational Research Association, San Francisco.

Hunt, E. (1995). Where and when to represent students this way and that way: An evaluation of approaches to diagnostic assessment. In P. Nichols, S. Chipman, & R. Brennan (Eds.*), Cognitively diagnostic assessment.* Hillsdale, NJ: Erlbaum.

Kyllonen, P.C., Christal, R.E., Woltz, D.J., Shute, V.J., Tirre, W.C., & Chaiken, S. (1990*). Cognitive Abilities Measurement (CAM) Test Battery, Version 4.0.* Unpublished computerized test battery, Armstrong Laboratory, Brooks Air Force Base, TX.

Martinez, M.E., & Katz I.R. (1992). *Cognitive processing requirements of constructed figural response and multiple-choice items in architecture assessment.* Princeton, NJ: Educational Testing Service.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from person's responses and performances as scientific inquiry into score meaning. *American Psychologist,* 50 (9), 741-749.

Nichols, P. (1995). *Another source of error in reliability theory.* Paper presented at the 1995 Annual Meeting of the American Educational Research Association, San Francisco.

Nichols, P.D., Chipman, S.F., & Brennan, R.L. (Eds.). (1995). *Cognitively diagnostic assessment.* Hillsdale, NJ: Erlbaum.

Shute, V.J. (1995). SMART: Student modeling approach for responsive tutoring. *User Modeling and User-Adapted Interaction, 5,* 1-44.

Shute, V.J., & Gluck. K.A., (1994). *Stat Lady: Descriptive Statistics Module.* Unpublished computer program, Armstrong Laboratory, Brooks Air Force Base, Texas.

Swanson, D.B., Norman, G.R., & Linn, R.L. (1995). Performance-based assessment: Lessons from the health professions. *Educational Researcher, 24*(4), 5-24.

Tatsuoka, K.K. (1995). Architecture of knowledge structures and cognitive diagnosis: A statistical pattern classification approach. In P. Nichols, S. Chipman, & R. Brennan (Eds.*), Cognitively diagnostic assessment.* Hillsdale, NJ: Erlbaum.

Moody, J. W. (1996). A multilevel modeling study of opportunity to learn and science achievement. Paper presented at the 1996 Annual Meeting of the American Educational Research Association, New York.

Yepes-Baraya, M., & Allen, N.L. (In press). *The process of identifying item attributes related to item performance for the 1993 National Assessment of Educational Progress (NAEP) science field test.* Princeton, NJ: Educational Testing Service.

MECHANICAL SPECIALTIES IN THE U.S. AIR FORCE: ACCESSION QUALITY AND SELECTION TEST VALIDITY

Stephen A. Truhon
Associate Professor
Department of Social Sciences

Winston-Salem State University
601 Martin Luther King Jr. Drive
Winston-Salem, NC 27110

Final Report for:
Summer Faculty Research Program
Armstrong Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, DC

and

Armstrong Laboratory
Brooks Air Force Base, TX

July 1996

MECHANICAL SPECIALTIES IN THE U.S. AIR FORCE: ACCESSION QUALITY AND SELECTION TEST VALIDITY

Stephen A. Truhon
Associate Professor
Department of Social Sciences
Winston-Salem State University

Abstract

The current study examined whether there has been a decline in mechanical abilities among airmen in Mechanical Air Force specialties (AFSs) and whether the Mechanical portion of the Armed Services Vocational Aptitude Battery (ASVAB) is a valid predictor of mechanical performance among those specialties. The records of 48,009 first-term recruits who enlisted in the service between January, 1990 and September, 1995 and were assigned to a Mechanical AFS were examined. Results indicated that the level of mechanical performance among those recruits selected for Mechanical specialties has remained stable. The Mechanical portion of the ASVAB appears to be a valid predictor of performance during technical school training. An explanation for these findings is discussed and other factors to improve the prediction of mechanical performance are considered.

MECHANICAL SPECIALTIES IN THE U.S. AIR FORCE:
ACCESSION QUALITY AND SELECTION TEST VALIDITY

Stephen A. Truhon

Winston-Salem State University

The Air Force, like the other American military organizations, uses the Armed Services Vocational Aptitude Battery (ASVAB) to select applicants for military service and to assign accepted applicants to specific jobs (Air Force Specialties [AFSs]). The ASVAB is made up of 10 subtests, which measure aptitudes in verbal, mathematical, clerical-speed, and technical areas. The Air Force uses four composites from the ASVAB to assign recruits to jobs in specialties: Mechanical, Administrative, General, and Electronic; and a fifth composite, the Armed Forces Qualification Test (AFQT), to select applicants for entry into the Air Force, prior to job assignment.

A number of changes occurring in the military, in general, and in Mechanical specialties, in particular, have created concern among recruiters. With the end of the Cold War, the amount of money available for military budgets has declined (Grier, 1995). The resulting drawdown has led to a smaller military force. A smaller force could work to the advantage of the military as it allows the Air Force to be more selective in the applicants it chooses. However, this drawdown has led some in the civilian population to believe that the military does not need new recruits and that it is not a stable career option. Surveys indicate the percentage of 16 to 21 year olds males interested in enlisting has been (Chapman, 1996).

Meanwhile other changes have particularly affected Mechanical specialties. While the number of commissioned officers in the Air Force (often pilots) has remained steady in recent years, the number of enlisted recruits (from which those in Mechanical AFSs are often selected) has been declining (Chapman, 1996). At the same time, increased demands are being made of mechanics. In 1992 as part of its Year of Training initiative, the Air Force began the Mission Ready Technician (MRT) program (Rankin, 1995). As part of MRT, airmen receive classroom and controlled technical training and then are sent to their assignments to receive operational training (Kuhn, 1995). Thus graduates of technical training in Mechanical AFSs are made productive members of their units much sooner than in the past. In addition there has been a

decline in mechanical skills among applicants to the Air Force (Skinner, 1996, unpublished data). Similar trends have been observed by other branches of the military (Defense Manpower Data Center, 1996, unpublished data) and by civilian industries (e.g., Orenstein, November 20, 1995).

Another concern has been the validity of the ASVAB in predicting performance in Mechanical AFSs. Studies from those recruits tested in the 1980s suggested that the ASVAB had acceptable validity. Wilbourn, Valentine & Ree (1984) reported a median uncorrected correlation between Mechanical score and final school grade of .41. Similarly, Ree and Earles (1992) reported a weighted mean uncorrected validity of .43 (.73, when corrected for restriction in range). Carey (1994) reported that, among Marine Corps automotive and helicopter mechanics, the ASVAB and time in the service accounted for a weighted mean uncorrected validity of .52 (.68, when corrected for restriction in range). In general, the ASVAB predicts well performance on Mechanical specialties.

It is important in these studies to correct for restriction in range, since only recruits with a sufficiently high score will be selected for military training. Because of this restriction in the scores on the ASVAB, the validity of the ASVAB for training is probably underestimated. Procedures to correct for range restriction are often parts of routines for conducting meta-analysis (Hunter & Schmidt, 1990).

These studies also suggest that there is room for improvement in predicting performance in Mechanical specialties. Among the tests suggested for use are: the AFQT (Ree & Earles, 1992; Wilbourn et al., 1984); the Electronics Aptitude Index from the ASVAB (Ree & Earles, 1992), and an object assembly test (Carey, 1994).

The current study is concerned with three questions: 1) has the decline in mechanical abilities among applicants affected the quality of Air Force recruits in Mechanical specialties?; 2) given the changes in the backgrounds of applicants, has the ASVAB retained its value as a predictor of training outcomes for Mechanical specialties?; and 3) what other factors could improve prediction of mechanical performance?

METHODS

Subjects

The subjects were 48,009 first-term recruits who enlisted in the service between January, 1990 and September, 1995 and were assigned to a Mechanical AFS. These recruits were primarily male, white, single, and had completed high school.

Measures

A database on the recruits was created from their PACE, MEPS, and technical school training records. In all, 155 variables were gathered. Of primary interest were the recruits' Mechanical score on the ASVAB (MECH) and their final school grade (FSG) in technical school. The Mechanical score is a percentile score composite from the Mechanical Comprehension (MC), General Science (GS), and Auto and Shop Information (AS) subtests of the ASVAB (with $\overline{X}$= 50 and s= 10). Ree and Earles (1992) reported that MECH has an internal consistency reliability of .90.

FSG is calculated by averaging the percent correct scores from a series of multiple-choice tests the recruit completes during technical training. For purposes of this study FSG is considered to have a reliability of .80 (Pearlman, Schmidt, & Hunter, 1980).

Criterion Groups

Sixty-two Mechanical AFSs were considered for examination. Instead of considering each of these AFSs as a separate sample, up to four samples were examined from each AFS. Interviews of training course managers of the recruits identified changes in course content, emphasis, length, location and performance approach. These changes was used to form new samples. A total of 144 possible samples were considered in the current study, but because some of the changes took place after these recruits began training, only 113 samples were included in this study (see Table 1).

TABLE 1

Mechanical Specialties Examined

Sample #[1]	New AFSC	Old AFSC	Graduation Dates	Description
1	2A3X3A	452X4A	before 9304	Tactical Aircraft Maintenance F-15
2			between 9304 and 9512	
4	2A3X3B	452XB	before 9305	Tactical Aircraft Maintenance F-16/F-117
5			between 9305 and 9504	
6			after 9504	
7	2A3X3C	452X4C	before 9402	Tactical Aircraft Maintenance F/EF-111
8			after 9402	
9	2A3X3E	452X4E	before 9407	Tactical Aircraft Maintenance A-10
10			between 9407 and 9602	
13	2A3X3H		after 9402	Tactical Aircraft Maintenance U-2
15	2A5X1A	457X2C	between 9304 and 9412	Aerospace Maintenance C-9/C-20/C-21/C-141/T-39/T-43
16			after 9412	
17	2A5X1B	457X2A	before 9412	Aerospace Maintenance C-12/C-26/C-27/C-130
18			between 9412 and 9512	
20	2A5X1C	457X2B	before 9406	Aerospace Maintenance C-5
21			between 9406 and 9605	
23	2A5X1D	457X2E	all	Aerospace Maintenance C-17
24	2A5X1E	457X0A	before 9401	Aerospace Maintenance B-1/B-2
25			between 9401 and 9601	
28	2A5X1G	457X0C	between 9401 and 9407	Aerospace Maintenance C-18/C-135/E-3/VC-25/VC-137
29			after 9407	
30	2A5X1H	457X0D	before 9312	Aerospace Maintenance KC-10/E-4
31			between 9312 and 9407	
32			after 9407	
33	2A5X2	457X1	before 9406	Helicopter Maintenance
34			after 9406	
35	2A6X1B	454X0B	before 9305	Aerospace Propulsion Turbo
36			after 9305	
37	2A6X1A	454X0A	before 9305	Aerospace Propulsion Jet
38			between 9305 and 9511	
41	2A6X2	454X1	before 9304	Aerospace Ground Equipment
42			between 9405 and 9509	
43			after 9509	
44	2A6X3	454X2	before 9307	Aircrew Egress Systems
45			after 9307	

[1] Some changes in course content, etc. took place after these recruits began training. As a result, some samples had no recruits in them and are not included for analysis.

TABLE 1
(continued)

Mechanical Specialties Examined

Sample #	New AFSC	Old AFSC	Graduation Dates	Description
46	2A6X4	454X3	before 9206	Aircraft Fuel Systems
47			between 9206 and 9504	
48			after 9504	
49	2A6X5	454X4	before 9304	Aircraft Hydraulics System
50			after 9304	
52	2A6X6		after 9304	Aircraft Electrical & Environmental Systems
53		454x5	before 9304	Strategic Electrical & Environmental Systems
54			after 9304	
55		452x5	before 9304	Tactical Electrical & Environmental Systems
56			after 9304	
57		454x6	before 9304	Airlift Electrical & Environmental Systems
58			after 9304	
59	2A7X1	458X0	before 9304	Aircraft Metals Technology
60			after 9304	
61	2A7X3	458X2	before 9302	Aircraft Structural Maintenance
62			between 9302 and 9501	
63			after 9501	
64	2A7X4	458X3	before 9309	Fabrication & Parachute
65			between 9309 and 9512	
67	2E6X1	361X0	before 9511	Communications Antenna Systems
69	2E6X2	361X1	before 9511	Communications Cable Systems
71	2F0X1	631X0	before 9205	Fuels
72			between 9205 and 9408	
73			after 9408	
74	2M0X2A	411X1A	before 9304	Missile & Space Systems Maintenance
75			between 9304 and 9406	
76			after 9408	
77	2T3X1	472X0	before 9305	Special Vehicle & Equipment Maintenance
78			between 9306 and 9612	
80	2T3X2A	472X1A	before 9305	Special Vehicle Maintenance Firetrucks
81			after 9305	
82	2T3X2B	472X1B	before 9305	Special Vehicle Maintenance Refueling Vehicles
84	2T4X1	472X2	before 9305	General Purpose Vehicle Maintenance
85			after 9305	
87	2T4X2	472X3	after 9306	Vehicle Body Maintenance
88	2T2X1	605X5	before 9305	Air Transportation
89			after 9305	

TABLE 1
(continued)

Mechanical Specialties Examined

Sample #	New AFSC	Old AFSC	Graduation Dates	Description
91	2W0X1A	465X0	between 9404 and 9603	Munitions Systems Material
93	2W0X1B	461X0	before 9404	Munitions Systems Production
94			between 9404 and 9603	
96	2W1X1C	462X0C	before 9404	Aircraft Armament Systems A-10
97			between 9404 and 9511	
99	2W1X1E	462X0E	before 9404	Aircraft Armament Systems F-15
100			between 9404 and 9511	
101			between 9511 and 9601	
103	2W1X1F	462X0F	before 9404	Aircraft Armament Systems F-16
104			between 9409 and 9511	
105			between 9511 and 9601	
107	2W1X1H	462XOH	before 9404	Aircraft Armament Systems F-111
109	2W1X1K	462X0K	before 9404	Aircraft Armament Systems B-52
110			between 9404 and 9511	
113	2W1X1L	462X0L	before 9404	Aircraft Armament Systems B-1
114			between 9404 and 9511	
116	2W1X1Z	462X0Z	before 9404	Aircraft Armament Systems All Other
117			after 9404	
118	2W2X1	463X0	before 9404	Nuclear Weapons
119			after 9404	
120	3E0X2	542X2	before 9405	Electric Power Production
121			after 9405	
122	3E1X1		before 9406	Heating, Ventilation, Air Conditioning & Refrigeration
123			after 9406	
124		545X0	all	Refrigeration & Air Conditioning
125		545X2	all	Heating Systems
126	3E2X1		all	Pavements & Construction
127		551x1	all	Masonry
128		551X0	all	Pavements Maintenance
129	3E3X1		before 9405	Structural
130			after 9405	
131		552X0	all	Carpentry
132	3E4X1		before 9405	Utilities Systems
133			after 9405	
134		552X2	all	Metals Fabricating
135		566X1	all	Utilities Systems
136		552X5	all	Plumbing
137	3E4X2	566X2	before 9405	Liquid Fuel Systems
138			between 9405 and 9501	
139	3E8X1	464X0	all	Explosive Ordnance Disposal

TABLE 1
(continued)

Mechanical Specialties Examined

Sample #	New AFSC	Old AFSC	Graduation Dates	Description
140	3P1X1	753X0	all	Combat Arms Training & Maintenance

Analysis

Analyses of variance were performed to determine changes in test performance over time. Correlational analyses were conducted to study the validity of the ASVAB, with meta-analyses used to correct for unreliability and range restriction.

RESULTS

Changes in Mechanical Ability Over Time

The first question of concern was whether the decline in mechanical abilities seen in the applicant population affects those in the Mechanical specialties. Visual inspection of the means in Table 2 suggest that there is no overall pattern. However, an analysis of variance reveals significant differences between the means of the Mechanical scores ($F(5, 46788) = 11.15$, $p < .0001$). Examination of the mean scores on the subtests that make up the Mechanical score shows significant effects among year of entry groups[2] for each of the tests ($F(5,46788) = 107.49$ (GS); $= 15.53$ (MC); and $= 22.00$ (AS)). Table 3 reveals that the GS and MC scores are increasing while the AS scores are decreasing. However, it should be noted that with the large number of recruits included ($N = 46{,}794$), the analysis of variance is quite powerful. Small differences in mean scores may be statistically significant, but of little practical value. Comparison of the year of entry means with the overall means reveal changes of usually less than .1 s.d. Likewise, the year of entry into the military accounts for a small proportion of the variance ($\eta^2 = .0012$

[2] Year of entry analyses for the MECH score are based on accessions for the entire calendar year (January-December) in 1990-1994 but for the January - September period in 1995. The same applies for analysis by year of entry described later in the paper.

for the Mechanical scores; η^2 = .0114 for the GS; η^2 = .0017 for MC; and η^2 = .0023 for AS).

TABLE 2

Comparison of Mean MECH, GS, MC, AS, and AFQT Scores by Year of Entry

Year of Entry	N	MECH Mean	S.D.	GS Mean	S.D.	MC Mean	S.D.	AS Mean	S.D.	AFQT Mean	S.D.
1990	8,591	72.53	14.23	18.56	3.09	18.36	3.14	18.16	3.83	62.12	14.66
1991	8,112	72.67	14.35	18.62	3.10	18.38	3.17	18.18	3.83	63.17	15.49
1992	10,401	72.60	14.27	18.61	3.12	18.50	3.09	18.07	3.79	62.61	15.78
1993	7,137	71.32	14.79	18.61	3.10	18.32	3.22	17.79	3.81	61.53	15.62
1994	8,190	72.25	13.75	19.34	2.91	18.63	3.24	17.75	3.65	65.01	15.57
1995	4,363	71.63	14.24	19.34	3.00	18.59	3.32	17.79	3.53	64.32	15.92

Changes in AFQT Scores Over Time

How have AFQT scores changed during the same period? An analysis of variance was performed on AFQT scores by year of entry. There was a significant change over time (F (5, 46788) = 53.90, $p < .0001$, η^2 =.0057). Examination of Table 2 reveals no pattern to the change in scores.

The Relationship between MECH and AFQT

Having analyzed the patterns of change for MECH and AFQT, the relationship between these two variables was then examined. The correlation between MECH and AFQT was moderate and significant (r = .34, df = 46,792, $p <$ .001). Performing separate correlations by year of entry revealed that the relationship was fairly consistent across the years (1990: r = .31; 1991: r = .35; 1992: r = .32; 1993: r = .37; 1994: r = .36; and 1995: r = .39)

Validity of Mechanical Score in Predicting Final School Grade

The weighted average correlation between MECH scores and FSG across the 113 samples was .31 (n= 40,654). As can be seen in Table 3, most of the validity correlations are between .20 and .40 although they range from .06 to .89. (Most of the extreme values occur when the sample size is small). It is

also notable that most of the validities are quite similar in magnitude for samples derived from the same AFS Code (AFSC). (Compare Table 1 with Table 3). However, there are some interesting exceptions to this finding. For example, samples 71, 72, and 73 from AFSC 2F0X1 (Fuels) have validity coefficients of .07, .15, and .29, respectively.

TABLE 3

Correlation between MECH and FSG for 113 Samples

Sample Number	n	r	Sample Number	n	r
1	1487	.36	72	384	.15
2	379	.53	73	464	.29
4	1491	.37	74	457	.18
5	90	.38	75	5	.89
6	292	.38	76	39	.14
7	276	.38	77	294	.37
8	139	.42	78	433	.41
9	122	.39	80	114	.09
10	129	.28	81	40	.52
13	77	.43	82	149	.23
15	149	.59	84	634	.36
16	96	.47	85	178	.48
17	252	.48	87	5	.22
18	239	.36	88	1851	.18
20	85	.47	89	1044	.24
21	155	.45	91	411	.19
23	232	.44	93	2106	.35
24	23	.28	94	1019	.38
25	72	.43	96	166	.34
28	99	.36	97	17	.21
29	147	.59	99	663	.34
30	35	.41	100	293	.29
31	29	.42	101	27	.33
32	175	.45	103	905	.38
33	49	.56	104	289	.35
34	79	.38	105	6	.36
35	357	.44	107	189	.25
36	287	.34	109	204	.27
37	1207	.21	110	31	.39
38	755	.37	113	107	.49
41	1425	.34	114	20	.56
42	1296	.35	116	108	.28
43	30	.46	117	19	.46
44	102	.36	118	473	.19
45	209	.15	119	12	.75
46	306	.34	120	597	.36
47	560	.40	121	187	.26
48	159	.36	122	67	.45
49	857	.12	123	252	.41
50	534	.38	124	813	.34

TABLE 3
(continued)

Correlation between MECH and FSG by Sample

Sample Number	n	r	Sample Number	n	r
52	724	.28	125	234	.38
53	391	.26	126	291	.43
54	30	.33	127	711	.47
55	317	.19	128	262	.42
56	54	.06	129	63	.30
57	536	.12	130	218	.28
58	35	.48	131	509	.27
59	165	.36	132	47	.30
60	98	.51	133	211	.34
61	1217	.25	134	233	.34
62	266	.26	135	410	.28
63	182	.33	136	291	.34
64	307	.14	137	156	.30
65	73	.20	138	43	.39
67	201	.11	139	336	.20
69	606	.31	140	134	.09
71	1749	.07			

These separate analyses can be viewed as having been derived from separate samples from the same population and thus can be combined into a meta-analysis (Hunter & Schmidt, 1990). A meta-analysis was conducted with corrections for artifacts due to unreliability and range restriction. Estimates of reliability for the predictor and the criterion were described earlier (i.e., .90 for the MECH, .80 for FSG). The estimate of range restriction relative to the ASVAB norm group was determined by dividing the standard deviation for MECH from each of the samples by 26.28. The program Metaquik (Stauffer, 1996) was used to conduct the meta-analysis. When corrected for unreliability and range restriction, the correlation between MECH and FSG is estimated at .60.

Metaquik also tests whether the samples are homogeneous, i.e., whether the sample statistics could have been derived from the same population parameter. A lack of homogeneity suggests the presence of one or more moderator variables, i.e., variables that cause differences in the correlation between the two variables being examined in the meta-analysis. Hunter, Schmidt, and Jackson (1982) suggest two methods for determining whether there are possible moderator effects: the 75% rule and a chi-square approximation.

The 75% rule says if 75% of the total variance in the sample statistics can be accounted for by artifacts (e.g., unreliability and range restriction), then the studies are homogeneous.If the chi-square[3] value is not significant, then the studies are homogenous. The 75% rule generally has more power than the chi-square approximation (Sackett, Harris, & Orr, 1986).

In current analysis, 20% of the variance (18% when corrected for unreliability and range restriction) is accounted for by artifacts. In the chi-square test χ^2 = 571.45 (627.65 corrected) with df = 112 (in both cases, p < .01). Both tests suggest that the samples are heterogeneous, i.e., there appear to be moderators. The small amount of variance accounted for by artifacts should not be surprising because the reliabilities of the MECH and FSG are constant for each of the samples and the estimates of range restriction for the MECH are between .40 and .50 for most of the samples.

Explaining the Relationship between MECH and FSG

The usual method for testing possible moderator variables is to categorize the samples into subsets based on a potential moderator and then perform a meta-analysis for each subset (Hunter & Schmidt, 1990, pp. 292-293). With the limited amount of time available to complete this report, that approach was foregone. This is, however, an area for further exploration and possible moderator variables will be described in the Discussion section.

Instead of the typical meta-analysis it was decided to examine relationships between other variables and MECH and FSG. This approach is akin to using meta-analysis as a means to model-building (Borman, White, Pulakos, & Oppler, 1991; Viswesvaran & Ones, 1995).

Two variables seemed to be promising: the number of high school shop classes; and the number of high school physical and applied sciences courses. The high school shop classes variable (HANDSON, so called because these classes provide students with hands-on mechanical experience) was calculated by counting the number of such classes (i.e., electronics, radio repair, auto

---

[3] The chi-square approximation is calculated by $\chi^2_{K-1} = Ns_r^2 / \left(1-\bar{r}^2\right)^2$, where K is the number of samples, N is the total number of subjects across the samples, s_r^2 is the variance of the correlations in the sample, and $\bar{r}$ is the average of the correlations in the sample.

repair, hydraulics, industrial arts, and mechanics) that the recruit had completed. The physical and applied sciences courses variable (PHYSCI) was calculated by counting the number of such classes (i.e., physics, chemistry, general science, blueprint reading, and shop math) that the recruit had completed.

Analyses of variance were performed to examine the changes in HANDSON and PHYSCI. Both showed significant changes over time (HANDSON: F (5, 48003) = 45.76, $p < .0001$; PHYSCI: F (5, 48003) = 37.68, $p < .0001$), but examination of Table 4 suggests no obvious patterns. Percentages of variance accounted for by year of entry are both rather low (HANDSON: $\eta^2 = .0047$; PHYSCI: $\eta^2 = .0039$)

TABLE 4

Comparison of Mean HANDSON and PHYSCI by Year of Entry

Year of Entry	N	Mean HANDSON Score	S.D.	Mean PHYSCI Score	S.D.
1990	8,651	2.27	1.62	1.91	1.07
1991	8,173	2.35	1.63	2.04	1.13
1992	10,503	2.39	1.66	2.04	1.11
1993	7,384	2.06	1.62	2.00	1.09
1994	8,640	2.17	1.67	2.08	1.15
1995	4,658	2.32	1.68	2.16	1.15

The correlations among MECH, FSG, HANDSON, and PHYSCI were then calculated for the 113 samples. Meta-analyses were conducted to determine the weighted mean correlations. The meta-analyses also corrected for unreliability. Reliabilities for HANDSON and PHYSCI were determined by calculating the internal consistencies of the scales for each of the 113 samples. The reliability of HANDSON was generally higher than of PHYSCI (average r's =.67 and .32 respectively). Their restrictions in range were unknown. The results are presented in Table 5.

HANDSON correlates strongly with MECH (uncorrected $r = .31$, corrected $r = .48$), but does not correlate well with FSG (uncorrected $r = .07$, corrected $r = .10$). PHYSCI has a somewhat lower correlation with MECH (uncorrected $r = .20$, corrected $r = .31$) but its correlation with FSG is somewhat higher (uncorrected $r = .16$, corrected $r = .17$). Finally, HANDSON and PHYSCI are moderately correlated (uncorrected $r = .32$, corrected $r = .71$).

TABLE 5

Correlations among MECH, FSG, HANDSON, and PHYSCI[4]

	MECH	FSG	HANDSON	PHYSCI
MECH		.60	.48	.31
FSG	.31		.10	.17
HANDSON	.37	.07		.71
PHYSCI	.20	.16	.32	

DISCUSSION

The current study was concerned with changes that have been occurring in mechanical abilities and in Mechanical AFSs in the 1990s. The results suggest that some of the worries about Mechanical AFSs are unfounded.

While there have been reports of a decline in mechanical abilities among applicants (Skinner, unpublished data), this decline does not seem to have affected the quality of airman selected for assignments in Mechanical AFSs. While there are changes in the mean MECH and AFQT scores for those in Mechanical AFSs during the 1990s, the overall pattern is not one of decline. Even among the ASVAB subtests that are used to form the composite MECH score, two of the subtests (GS and MC) show increases, while the decrease in AS is small. While the decline in the number of applicants and the mechanical ability of those applicants have created difficulties for recruiters (Chapman, 1996), they still seem to be attracting a high quality of recruits for Mechanical specialties. This contention is further supported by the fact that the average number of shop and physical science classes taken by airman in Mechanical specialties was similar across year of entry groups.

Hypotheses about the causes of the decline in mechanical ability among accessions have generally focused on the changing nature of accessions (i.e., gender and racial composition). Yet the gender composition of accessions in Mechanical specialties do not appear to have changed appreciably in the 1990s and the racial composition has only changes slightly. Follow-on analyses

4 The correlations above the diagonal are corrected for unreliability and, in the relationship between MECH and FSG, for range restriction. The correlations below the diagonal are the weighted average of correlations from the samples.

revealed that, from 1990 to 1995, the percentage of females ranged from 4.2 to 6.0%, of blacks from 6.4 to 8.7%, and of other races from 2.7 to 7.3%.

Another hypothesis is that the Air Force, in seeking recruits with higher AFQT scores, has made it more difficult to obtain recruits with good mechanical skills. However, AFQT and MECH scores are positively and moderately correlated in the current study. Studies with a less restricted sample should have higher correlations.

The current study also replicated ASVAB validity studies done in the 1980s (Ree & Earles, 1992; Wilbourn et al., 1984) and demonstrated the validity of the ASVAB in predicting performance by recruits in Mechanical AFSs in the 1990s. All of the validity coefficients are positive, meaning that recruits with higher MECH scores were expected to demonstrate higher levels of academic performance on entry-level material taught during technical training.

It should be noted that the mean validity coefficients in the current study (uncorrected r = 32, corrected r = .60) are somewhat lower than those found in previous studies (generally, uncorrected r's$\approx$ .40, corrected r's $\approx$ .70) One reason for this difference may be in the ways in the studies were conducted. Both Ree & Earles (1992) and Wilbourn et al. (1984) conducted their analyses at the level of the AFS. In the current study, analyses were performed for groups within AFSs to account for changes in course content, emphasis, length, location and performance approach. It is also possible that changes in selection procedures may result in differences between the samples. Finally there may be changes in the nature of the Mechanical AFSs themselves. In studying recruits in the 1980s, Ree and Earles (1992) suggested that the Electronics score may be a better predictor of technical school training for some Mechanical AFSs than MECH score. This may be even truer today as working with machines requires additional electronic and computer knowledge.

Time constraints prevented this study from examining possible moderator variables between MECH and FSG. Clearly this is an area for future research. Potential moderator variables include: 1) level of aptitude requirement (i.e., how high a MECH score is required for entry into a particular AFS); 2) the type of aptitude required (i.e., whether the AFS requires a specific score on the MECH alone, on the MECH and on the Electronics composites, or on either the MECH or the Electronics composites); 3) type of career field (either determined by the first two digits of the AFSC or the clusters described by

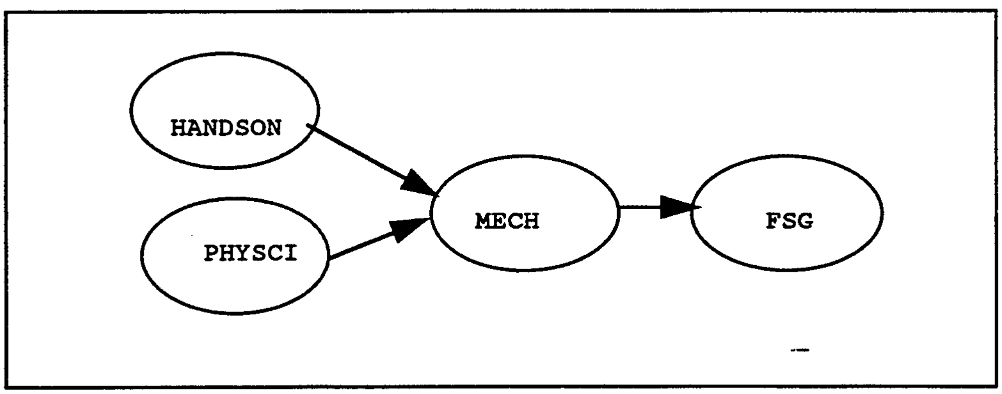
Alley & Ree (in preparation); and 4) an indicator of the nature of the training (e.g., the time or location of the training).

Viswesvaran and Ones (1995) present a strong case for the use of meta-analyses in the process of model building. They suggest procedures for using the estimated true correlations obtained through meta-analysis as input for structural equations modeling. Future analyses could follow this approach in helping to explain the relationship between MECH score and FSG.

The number of high school shop and applied science classes appear to be salient factors. The variables derived from them (HANDSON and PHYSCI) correlate well with each other. HANDSON correlates fairly strongly with MECH but not FSG. PHYSCI follows the same pattern but with lower correlations. These results suggest a basic model in which HANDSON and PHYSCI have a direct effect on MECH and an indirect effect on FSG, such as seen in Figure 1.

FIGURE 1

Basic Model Concerning the Relationships among HANDSON, PHYSCI, MECH, and FSG



The basic model is open to improvements. Improvements could be made to the measurement of HANDSON and PHYSCI by obtaining estimates of range restriction. In addition a variable counting the number of mathematics courses taken in high school (MATH) could be added. Gender differences have been noted in the number of females in mechanical AFSs, in their scores on several of these measures, as well as in the curriculum they pursue in high school. Such a model should also include gender. Instead of using the MECH score itself, the subtests (MC, GS, and AS) that make up MECH and the other subtests of the ASVAB (Arithmetic Reasoning [AR], Word Knowledge [WK], Paragraph Comprehension [PC], Numerical Operations [NO], Coding Speed [CS], Math Knowledge [MK], and

Electronic Information [EI]) could be used in a refined model. An index of training performance could be added for those AFSs for which MRT is applicable. Indicators of job performance in various job types within individual Mechanical AFSs after completion of entry-level training (e.g., six months, a year, or as long as four to eight years after training) could be used to illustrate how level of achievement in training, as measured by FSG, is related to later performance of technical tasks on the job. These suggestions and refinements are show in the more comprehensive model in Figure 2. This model depicts relatively simple and direct relationships among educational, aptitude, training, job assignment, and job performance. Follow-up research exploring the accuracy of alternate models is needed. As part of the research, the potential for using occupational surveys to explore measures such as the number, type, and difficulty of tasks performed for capturing the productivity of airmen is recommended.

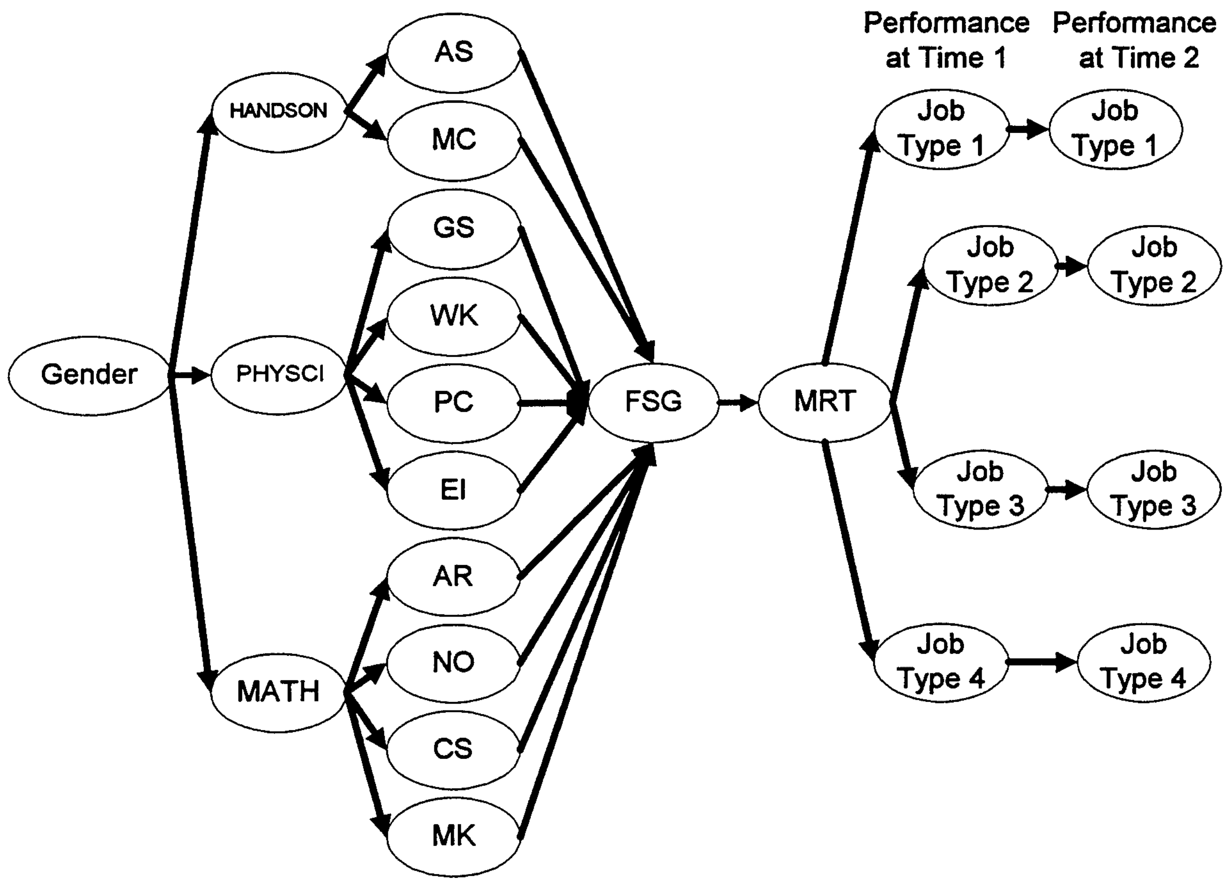
In conclusion, the current study found that 1) recruits in Mechanical AFSs show no decline mechanical ability; 2) the MECH composite of the ASVAB is a valid predictor of performance in technical school training; and 3) performance in certain high school courses may improve prediction of mechanical performance.

ACKNOWLEDGMENTS

I would like to thank the following persons for their assistance in this report: James R. Van Scotter, Jacobina Skinner, James Brazel, Melody Darby, James Earles, Patrick Kyllonen, Larry Looper, William Phalen, and Malcolm Ree.

FIGURE 2

Comprehensive Model of Relationships Among Educational Aptitude, Training, Job Assignment and Job Performance



REFERENCES

Alley, W.E., & Ree, M.J. (in preparation). Improved training performance through optimum personnel classification. (AL/HR-TP-XX-XX) Brooks AFB, TX: Human Resources Directorate, Armstrong Laboratory.

Borman, W.C., White, L.A., Pulakos, E.D., & Oppler, S.H. (1991). Models of supervisory job performance ratings. Journal of Applied Psychology, 76, 863-872.

Carey, N.B. (1994). Computer predictors of mechanical job performance: Marine Corps findings. Military Psychology, 6, 1-30.

Chapman, S. (1996). Uncertainty on the personnel front. Air Force Magazine, 79 (3), 40-43.

Earles, J.A. & Ree, M.J. (1992). The predictive validity of the ASVAB for training grades. Educational and Psychological Measurement, 52, 721-725.

Grier, P. (1995). Snapshots of a force on the move. Air Force Magazine, 78 (6), 58-62.

Hunter, J.E., & Schmidt, F.L. (1990). Methods of meta-analysis: Correcting error and bias in research findings. Newbury Park, CA: Sage Publications.

Hunter, J.E., Schmidt, F.L., & Jackson, G.B. (1982). Meta-analysis: Cumulating Research findings across studies. Beverly Hills, CA: Sage Publications.

Kuhn, T. (1995). Crew chiefs: Mission ready!. Airman, 39 (3), 30-35.

Orenstein, B.W. (November 20, 1995). Machinist shortage plagues manufacturers. Eastern Pennsylvania Business Journal.

Pearlman, K., Schmidt, F.L., & Hunter, J.E. (1980). Validity generalization results for tests used to predict job proficiency and training success in clerical occupations. Journal of Applied Psychology, 65, 373-406.

Rankin, K.S. (1995). The year of tech training. Airman, 39 (10), 38-41.

Ree, M.J. & Earles, J.A. (1992). Subtest and composite validity of ASVAB Forms 11, 12, and 13 for technical training courses (AFHRL-TR-1991-0107). Brooks AFB, TX: Human Resources Directorate, Manpower and Personnel Research Division.

Sackett, P.R., Harris, M.M., & Orr, J.M. (1986). On seeking moderator variables in the meta-analysis of correlational data: A Monte Carlo investigation of statistical power and type I error. Journal of Applied Psychology, 71, 302-310.

Stauffer, J.M. (1996). Metaquik 16: Psychometric meta-analysis program for Windows [Computer Program]. Terre Haute, IN: Department of Management and Finance, School of Business, Indiana State University.

Viswesvaran, C. & Ones, D.Z. (1995). Theory testing: Combining psychometric meta-analysis and structural equations modeling. Personnel Psychology, 48, 865-885.

Wilbourn, J.M., Valentine, Jr., L.D., & Ree, M.J. (1984). Relationships of the Armed Services Vocational Aptitude Battery (ASVAB) Forms 8, 9, 10 to Air Force technical school final grades (AFHR-TP-84-8). Brooks AFB, TX: Manpower and Personnel Division.

VALIDATION OF THE DEFORMABLE NECK MODEL FOR A +Gz ACCELERATION

Mariusz Ziejewski, Ph.D.
Associate Professor
Mechanical Engineering Department

North Dakota State University
Dolve 111
Fargo, ND 58105

Final Report for:
Summer Faculty Research Program
Armstrong Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, DC

and

Armstrong Laboratory

September 1996

VALIDATION OF THE DEFORMABLE NECK MODEL FOR A +Gz ACCELERATION

Mariusz Ziejewski, Ph.D.
Associate Professor
Mechanical Engineering Department
North Dakota State University

Abstract

During the ejection phase of escape, crew members are susceptible to neck injuries. Testing and computer simulations with the Articulated Total Body (ATB) computer program have been used to evaluate the effect of acceleration levels on human body response during ejection procedures. The objectives of this study were to create finite element neck models for several Vertical Drop Tower (VDT) test subjects to be incorporated into the deformable neck option of the ATB computer model and to assess the accuracy of the ATB model with the deformable neck option in predicting human response in the catapult phase of an ejection. The experimental data used in this study were collected from the biodynamic responses of human volunteers during an acceleration in the z-direction on the Vertical Drop Tower facility at Armstrong Laboratory at WPAFB. The experiments were performed for an approximate maximum acceleration of 10 G's for the male subjects and 8 G's for the female subject; the subjects were not wearing helmets. Data from twelve male and one female subjects were used for this study. A three segment model including the upper torso, neck and head was used. Head acceleration data at the mouth piece location and at the center of gravity location were calculated. The simulation results with the experimental data and, for reference purposes, the experimental chest acceleration data that were used as input into the ATB model were presented.

The ATB simulations using the current deformable neck option predict well the head acceleration in the x-direction which represents head rotation, however, they underestimate the maximum acceleration in the z-direction by up to 30%. Data from the analysis indicated that the location of the mouth piece on the subject is an important factor affecting the accuracy of simulation. The precise position of the test subject's head at the time of the impact could also affect the accuracy of simulation. It is anticipated that the neck load from the deformable neck option would provide a reasonable bending torque and would underestimate the compression load by 10 to 30 %.

VALIDATION OF THE DEFORMABLE NECK MODEL
FOR A +Gz ACCELERATION

Mariusz Ziejewski, Ph.D.

INTRODUCTION

During the ejection phase of escape, crew members are susceptible to neck related injuries. The United States Air Force initiated testing to evaluate the affect of acceleration levels to human neck response during ejection procedures. The main objective of their study was to define neck response during the catapult or impact acceleration phase of the ejection. Their follow-up objective was to define the specifications or criteria for allowable head mounted mass and center of gravity location that is safe for the crew members.

In addition to experimental efforts to evaluate the affect of acceleration levels on human neck response, an analytical computer program called Articulated Total Body (ATB) is used to simulate human response to these dynamic environments and hence used to estimate the human neck loads experienced during the catapult phase of ejection. A deformable neck option of the ATB model, based on an ANSYS finite element model, has been developed to more accurately predict human neck response. This option has been partially validated against frontal impact sled test results. However, it has not been validated against accelerations representing the catapult phase of ejection.

OBJECTIVES

1. To create finite element neck models for several Vertical Drop Tower (VDT) test subjects to be incorporated into the deformable neck option of the ATB computer model.
2. To assess the accuracy of ATB simulations with the deformable neck option in predicting human response in a catapult phase of an ejection.

EXPERIMENTAL DATA

Test Setup

The experimental data used in this study came from the Vertical Drop Tower study, VWI 199101. Male subjects came from Cell CA while the data for the female subject came from Cell BA. The experiments were performed for an approximate maximum acceleration of 10 G's for the male subjects, 8 G's for the female subject and the subjects were not wearing helmets.

The purpose of the VDT is to simulate an ejection seat catapult acceleration pulse by generating a +z-axis impact acceleration using a hydraulic decelerator. A generic seat with a restraint system which properly

positions the subject and a data acquisition system are mounted on a carriage which is positioned on the two vertical guide rails of the tower. The carriage is raised to a pre-determined height and then allowed to free fall into a water reservoir which acts as a hydraulic decelerator. A contoured piston mounted on the bottom of the carriage is guided into the water reservoir where the displacement of the water around the piston decelerates the carriage. Therefore, the drop height of the carriage and the shape of the piston control the magnitude and rise time of the acceleration pulse. During each test on the VDT, several channels of data are collected including acceleration and velocity of the drop carriage, linear and angular accelerations of the subject's head and chest, and forces in the seat, restraint system and subject (1). In order to collect the linear and angular accelerations of the head, a special bite bar (mouthpiece) instrumented with transducers is held in the subject's mouth.

Subjects

Thirteen subjects were used in this study, twelve male and one female. The data pertaining to their gender, weight, and height is given in Table 1.

Table 1 Test Subject Data

Test	Subject	Gender	Weight (N)	Height (cm)
2292	L7	Male	672	174.8
2293	T6	Male	854	174.4
2295	B1	Male	654	179.1
2297	B9	Male	663	173
2299	L9	Male	681	180.4
2301	L8	Male	827	179.1
2308	K5	Male	863	174.8
2309	C8	Male	778	184.2
2429	H11	Male	689	163.2
2442	F6	Male	836	177.8
2504	W7	Male	890	172.1
2505	G8	Male	734	174.6
2317	R13	Female	672	165.8

COMPUTER ANALYSIS

In this portion of the study two computer programs were used, the ANSYS finite element program and the Articulated Total Body (ATB) computer program with the newly developed deformable neck option. ANSYS is the linear finite element code that was used in this study to perform analysis of the deformable neck segment. The ATB model is a rigid body dynamics program used to predict the mechanical response of the human body in different dynamic environments such as aircraft pilot ejections, sled tests, etc. The deformable neck option of the ATB model is based on an ANSYS finite element model developed for the neck segments of human subjects to represent their deformability (2, 3).

The ANSYS finite element model has a cylindrical shape with inserts at the top and bottom for use in joint connection in the deformable neck option of the ATB model. The finite element model is used to determine the mode shapes and natural frequencies of the neck which are used in the deformable neck option of the ATB program to calculate the displacement of the neck.

For this analysis of human response in a simulated ejection acceleration pulse, ATB was setup as a three segment, two joint model, with segments representing the upper torso, the neck and the head, and with articulations at the upper torso/neck junction, and the neck/head junction. Experimental chest acceleration data was input into the ATB model as the acceleration of the chest segment of the three segment model.

In this study three main steps were taken, namely, determination of physical parameters of the subject's neck, generation of an ANSYS finite element model of the subject's neck, and performance of an ATB simulation. These steps were repeated for each subject included in this study.

Determination of Neck Parameters

In order to generate a finite element model of a particular subject, the length and radius of the neck must be obtained. In this study a computer program called Generator of Body Data (GEBOD) was utilized to calculate the length and weight of the neck segment. GEBOD is a computer program that generates human data sets for use in dynamic modeling (4). The type of data includes geometric and inertial properties of different body segments, and various joint locations and range of motion. For each subject, their gender, height and weight was entered into GEBOD. The output obtained after running the GEBOD program included the length and weight of the neck segment. The neck radius, determined based on these values, are given in Table 2.

Table 2 Test Subject Neck Data

Test	Subject	Gender	Neck Length (cm)	Neck Weight (N)	Neck Radius (cm)
2292	L7	Male	10.3	9.8	4.6
2293	T6	Male	9.4	11.3	5.4
2295	B1	Male	10.8	8.3	4.3
2297	B9	Male	10.2	8.8	4.6
2299	L9	Male	10.8	8.6	4.4
2301	L8	Male	9.9	10.7	5.1
2308	K5	Male	9.4	11.4	5.4
2309	C8	Male	10.6	9.8	4.7
2429	H11	Male	9.2	9.7	5.1
2442	F6	Male	9.8	10.9	5.2
2504	W7	Male	9	12	5.7

2505	G8	Male	10	9.7	4.9
2317	R13	Female	8.9	7.1	4.4

Generation of ANSYS Finite Element Model

Thirteen finite element models were created, one for each subject of the study using the corresponding neck length as provided by GEBOD and the radius as calculated. A Young's Modulus of 13.79 Mpa and a modal damping ratio of 0.3 were used in all thirteen of the finite element models (2).

The size of the finite element neck model is dependent on the length of and radius of the subject's neck and is therefore different for each subject. The number of nodes and elements for each model, along with the top and bottom node numbers used in joint connections with the ATB model are given in Table 3.

Table 3 Number of Nodes, Elements and Node Numbers Used in Joint Connections

Test	Subject	Gender	Number of Nodes	Number of Elements	Top Node	Bottom Node
2292	L7	Male	1114	752	441	552
2293	T6	Male	1506	1064	586	727
2295	B1	Male	1006	682	401	500
2297	B9	Male	1114	752	441	552
2299	L9	Male	1114	752	441	552
2301	L8	Male	1250	838	481	611
2308	K5	Male	1506	1064	586	727
2309	C8	Male	1114	752	441	552
2429	H11	Male	1162	772	433	563
2442	F6	Male	1250	838	481	611
2504	W7	Male	1506	1064	586	727
2505	G8	Male	1250	838	481	611
2317	R13	Female	1034	692	397	508

After the solution phase of ANSYS is completed, ANSYS contains all the information needed for the ATB input file. The first four vibration modes were extracted for each subject using a macro developed for this purpose (2). This macro collects data from the ANSYS output and places it in a data file. The data file is then converted with a FORTRAN program into a file that is used in the deformable neck option of ATB (2).

ATB Simulation

The ATB input file was modified to include the information obtained from the finite element analysis. The modifications included the name of the file containing the finite element data, the node number of the insertion nodes for both the top and bottom of the neck, and the experimental chest acceleration. The

results from the ATB simulations were then compared against the experimental data and are presented in the results section of this report.

RESULTS

Several preliminary ATB simulations were performed using selected modes as determined by ANSYS. The preliminary ATB simulations using the first two modes showed good agreement with experimental data. The addition of the third mode (torsional) did not change the results of the simulation. With the addition of the fourth mode (compression), excessive oscillations appeared as seen in Figure 1. Based on the preliminary analysis the first two mode shapes as determined by ANSYS were used in subsequent ATB simulations.

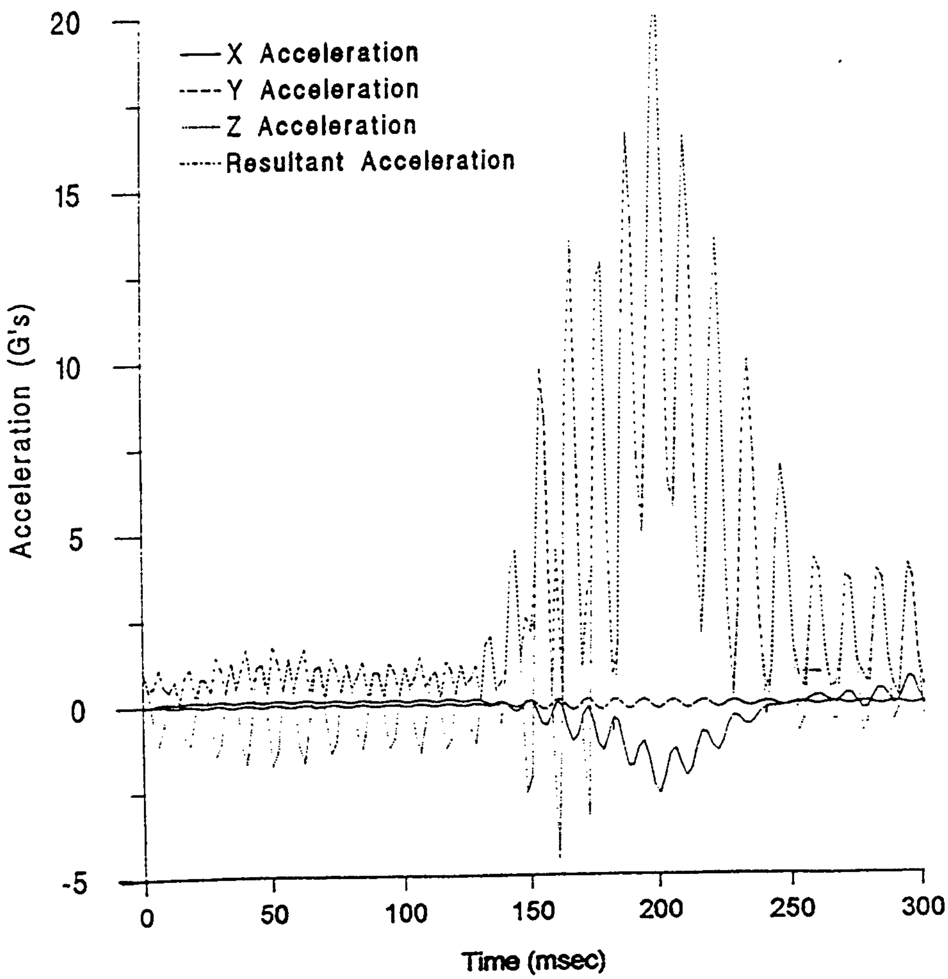


Figure 1 Head Acceleration Using 1st, 2nd and 4th Mode Shapes

Thirteen ATB simulations using the first two mode shapes were performed. Head acceleration data at the mouthpiece location and at the center of gravity of the head were calculated. The mouth piece location

with respect to the head local coordinate system at the center of mass was assumed to be the same for each subject. The simulation results along with the experimental data and for references purposes, the chest acceleration data for each subject are given in Figures 2 through 14.

In each figure, the left graph shows the simulated versus the experimental head acceleration at the mouthpiece. As can be seen, the maximum simulated acceleration in the z-direction always underestimates the experimental acceleration by 10 to 30 %. The simulated head acceleration in the x-direction at the mouth piece location shows agreement with experimental data. The acceleration in the x-direction is a representation of head rotation. The observed discrepancies between the simulated and experimental data in the x-direction could partially be due to the fact that the exact position of the test subject's head at the time of impact is not included.

The middle graph shows the simulated response at the center of gravity and the experimental data at the mouthpiece. The observed dissimilarities in the simulated results at the center of gravity and the experimental data at the mouth piece locations are mainly due to their relative positions from the center of head rotation.

Comparison between the acceleration curves at the mouth piece (given in the left graph) and at the center of gravity of the head (given in the middle graph) indicate that the location of the data collection point for the analysis has a significant influence on the outcome of the simulation.

For reference purposes, the right graph depicts the experimental chest acceleration in the z-direction.

CONCLUSIONS

1. The ATB simulation of the head acceleration using the current deformable neck option:
    a) predicts the general trends of the human head response
    b) predicts well the acceleration in the x-direction indicating good prediction of head rotation
    c) underestimates the maximum acceleration in the z-direction by 10 to 30%
2. The location of the mouth piece of the test subject is an important factor affecting the accuracy of simulation.
3. The position of the test subject's head at the time of the impact can be an important factor affecting the accuracy of simulation.
4. It is anticipated that the neck load from the deformable neck option would provide a reasonable bending torque and would underestimate the compression load by 10 to 30 %.

SUGGESTIONS FOR FURTHER RESEARCH

1. Since the neck model was originally developed to predict frontal impact response, further effort into properly representing the human neck response to compressive forces in the deformable option of ATB is needed.
2. In order to perform complete validation of the deformable neck option against all available experimental data, a comprehensive statistical analysis of this data, leading to determination of typical response characteristics as a function of anthropomorphic parameters is needed.
3. In order to determine typical response characteristics based on anthropomorphic parameters, the determination of those parameters that significantly influence the human body response must be identified.

REFERENCES

1. Perry, C.E. "Biomechanics of Ejection Safety for Helmet-Mounted Systems", Thesis submitted to Wright State University, 1995.
2. Ashrafiuon, H., Colbert, R., Obergefell, L., Kaleps, I. "Modification of the ATB Model to Incorporate Deformable Bodies and its Verification Through Dummy Neck Modeling and Simulation" Vulnerability and Assessment Branch, Crew Systems Directorate, Armstrong Laboratory, Wright-Patterson Air Force Base
3. Ashrafiuon, H., Colbert, R., Obergefell, L., Kaleps, I. "Finite Element Modeling of the Human Neck and Its Validation for the ATB" Vulnerability and Assessment Branch, Crew Systems Directorate, Armstrong Laboratory, Wright-Patterson Air Force Base
4. Cheng, H.; Obergefell, L.; Rizer, A. "Generator of Body Data (GEBOD) Manual", AL/CF-TR-1994-0051, Air Force Material Command, Wright-Patterson Air Force Base, Ohio

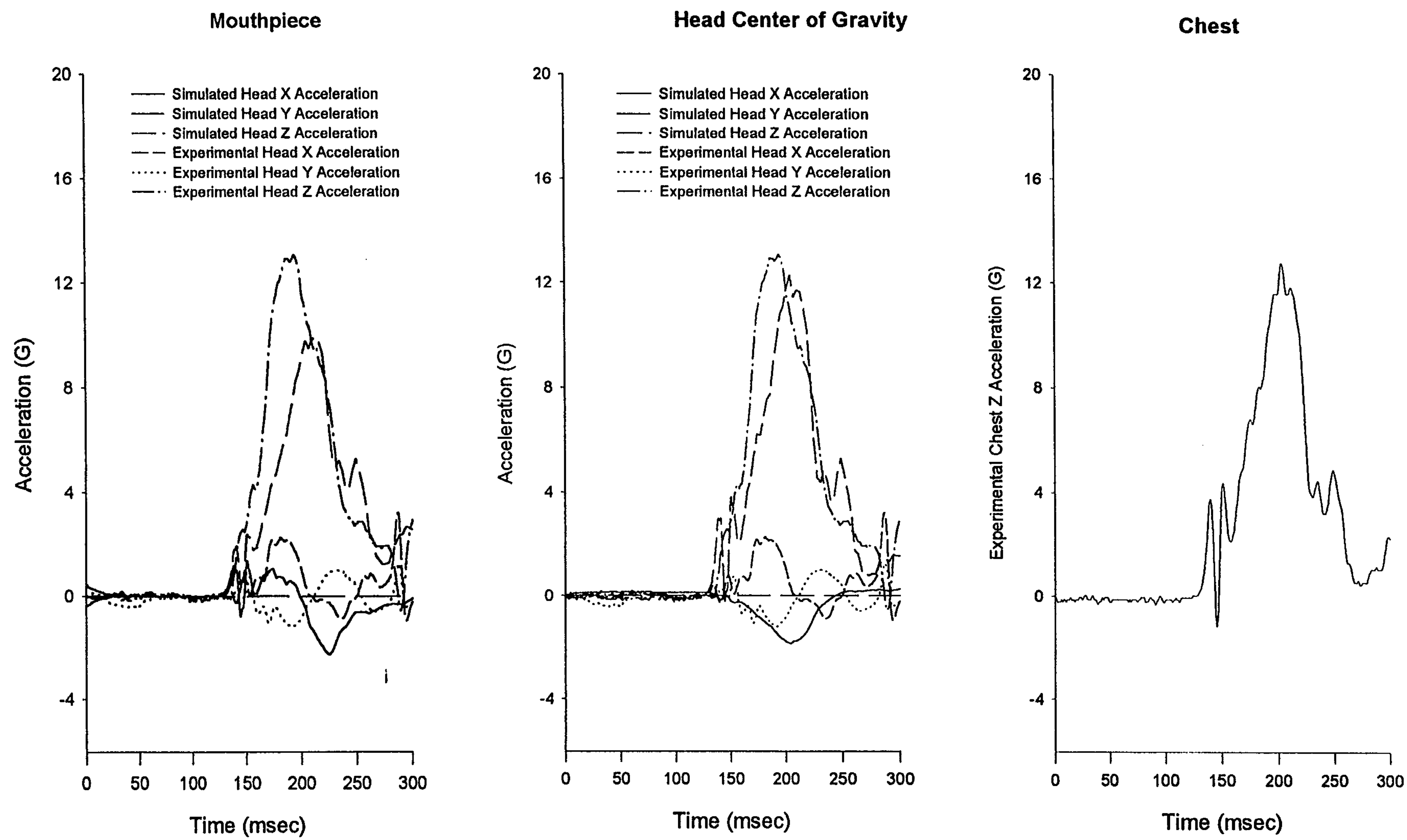


Figure 2 Acceleration data for mouthpiece, head center of gravity and chest for male subject 2292.

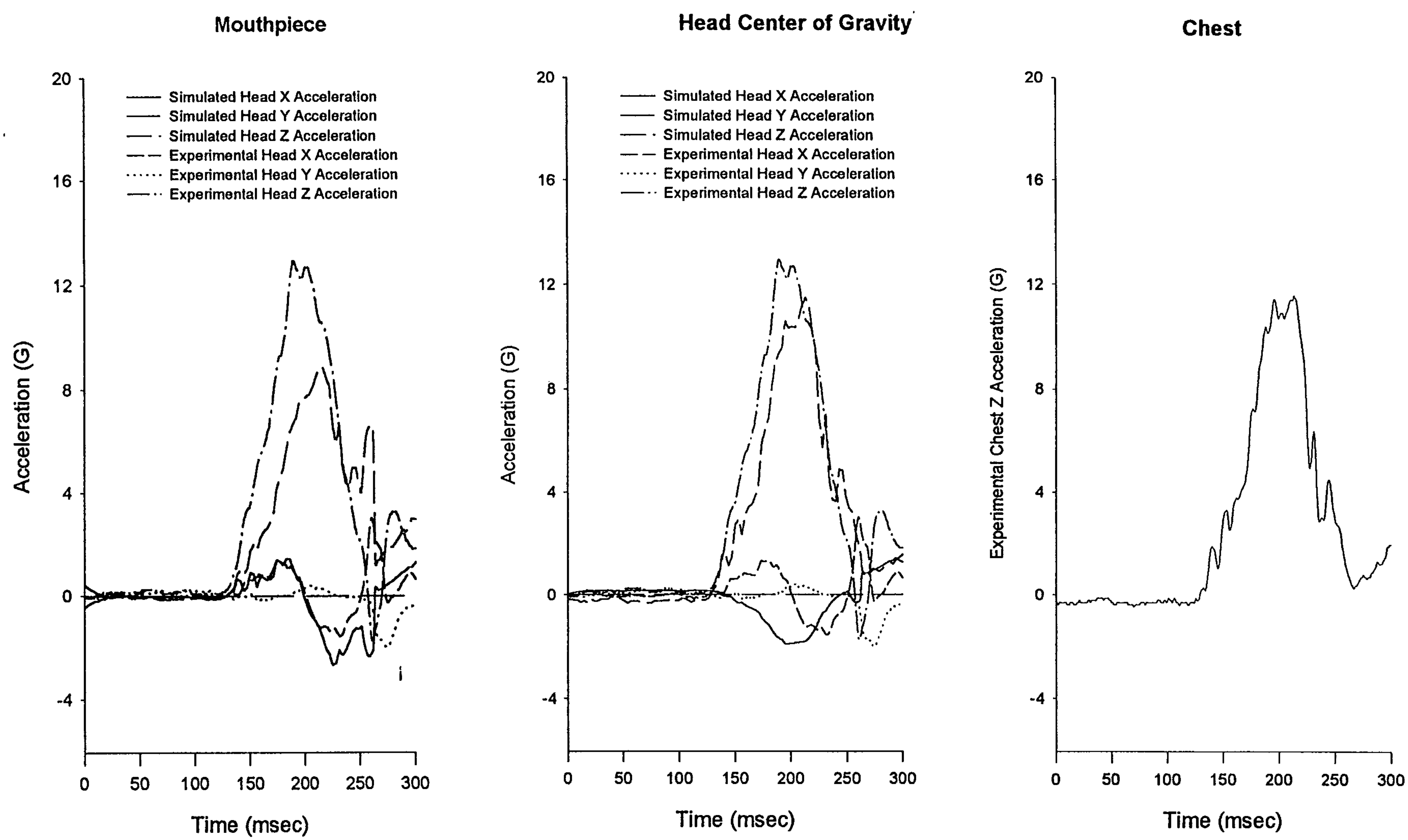


Figure 3 Acceleration data for mouthpiece, head center of gravity and chest for male subject 2293.

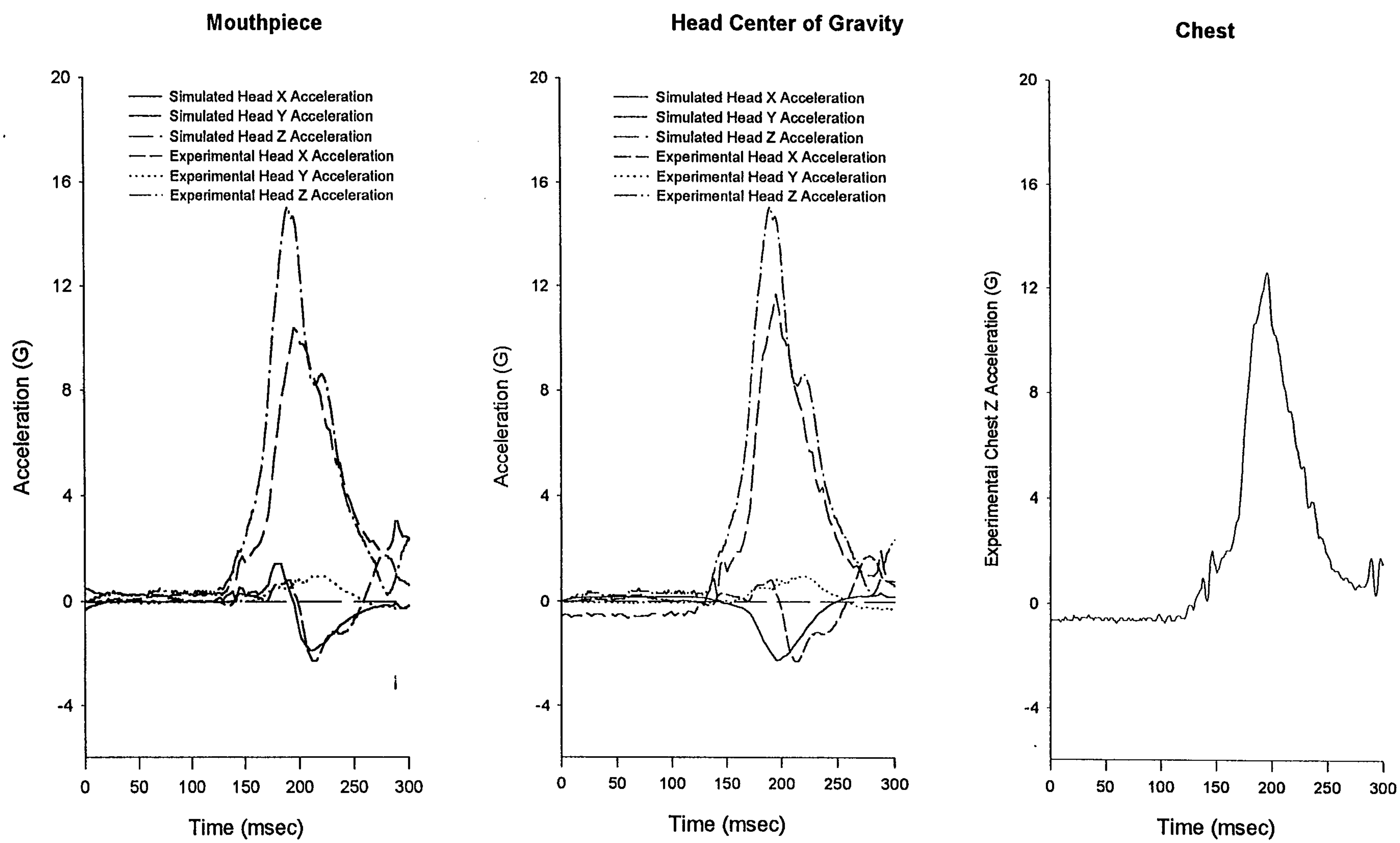


Figure 4 Acceleration data for mouthpiece, head center of gravity and chest for male subject 2295.

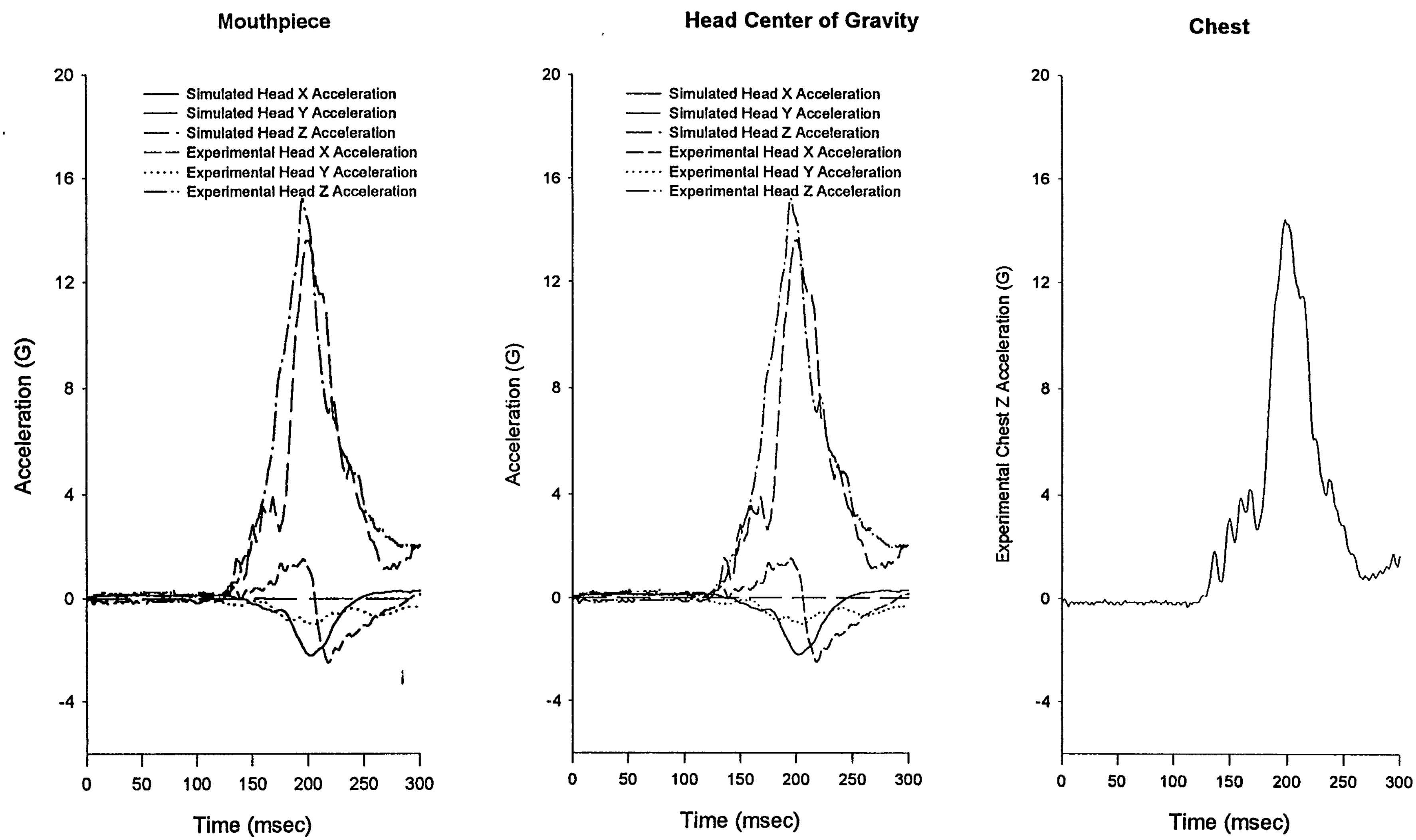


Figure 5 Acceleration data for mouthpiece, head center of gravity and chest for male subject 2297.

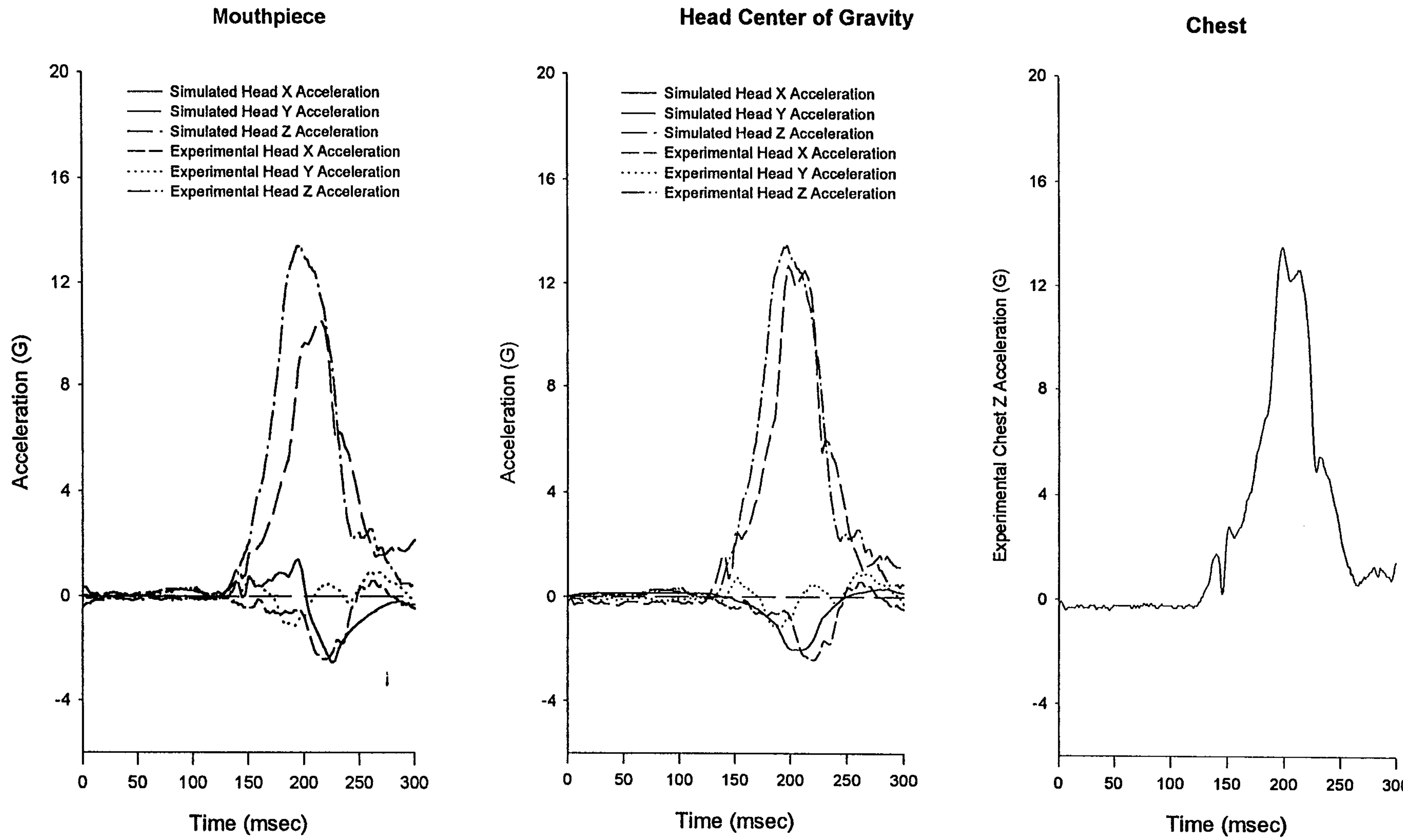


Figure 6 Acceleration data for mouthpiece, head center of gravity and chest for male subject 2299.

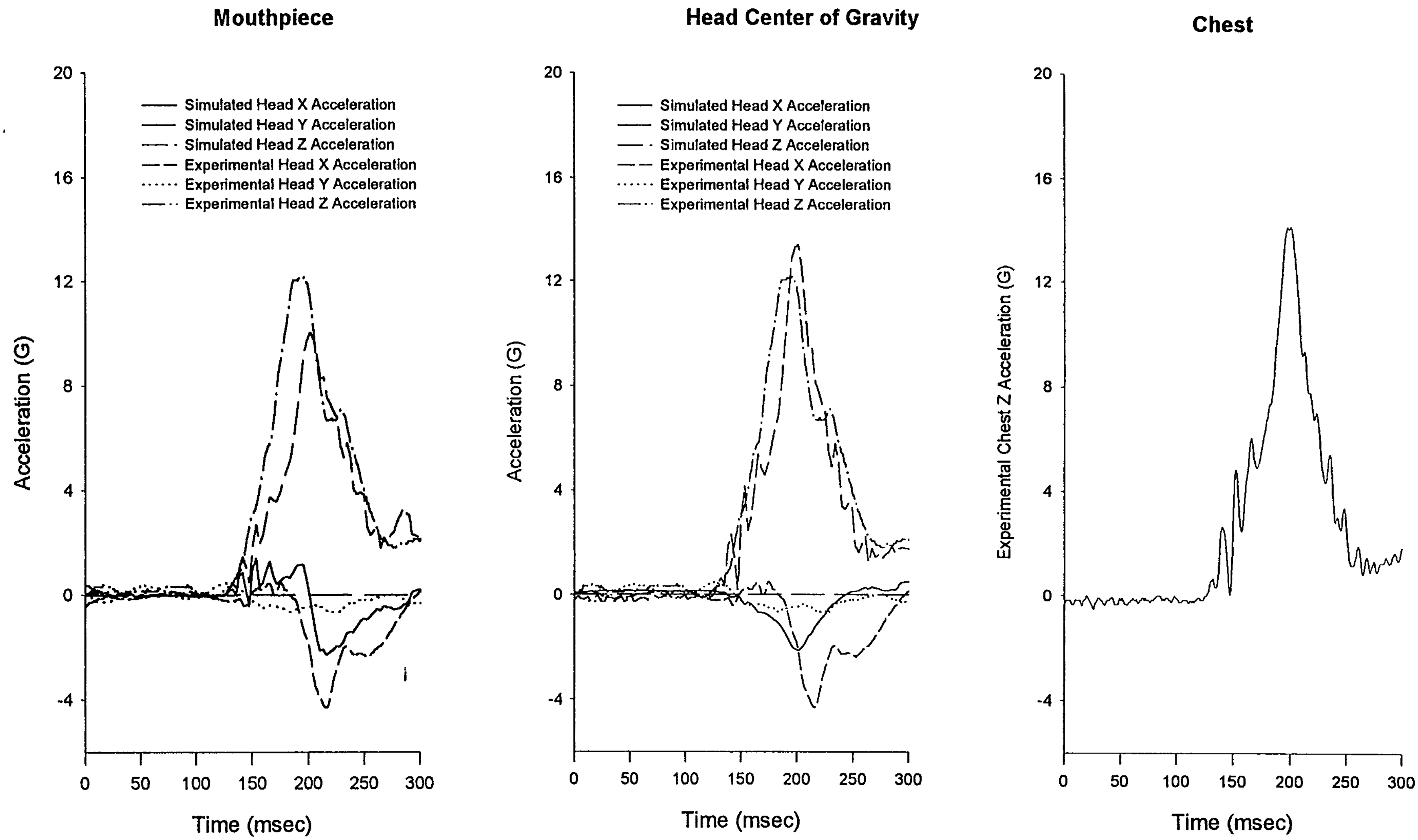


Figure 7 Acceleration data for mouthpiece, head center of gravity and chest for male subject 2301.

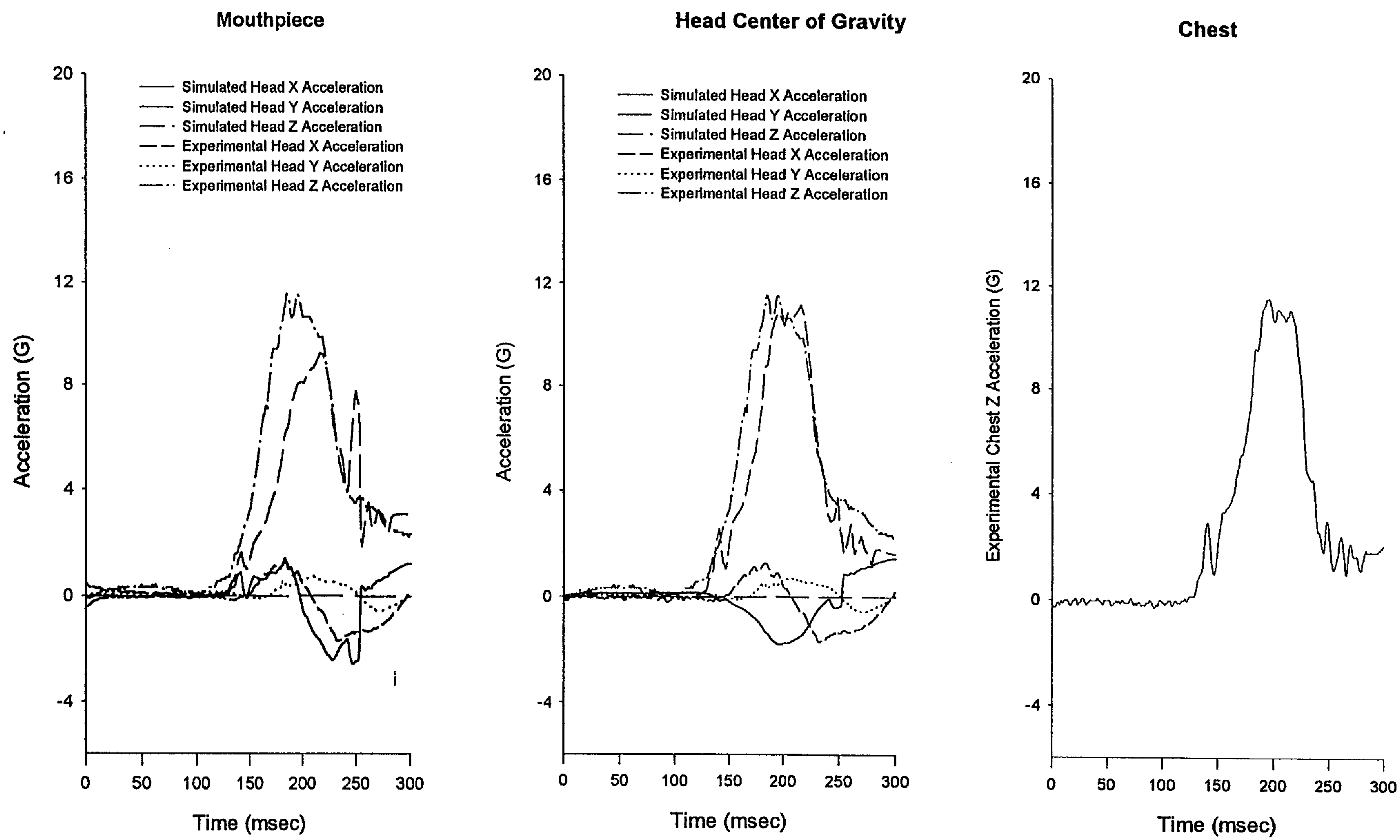


Figure 8 Acceleration data for mouthpiece, head center of gravity and chest for male subject 2308.

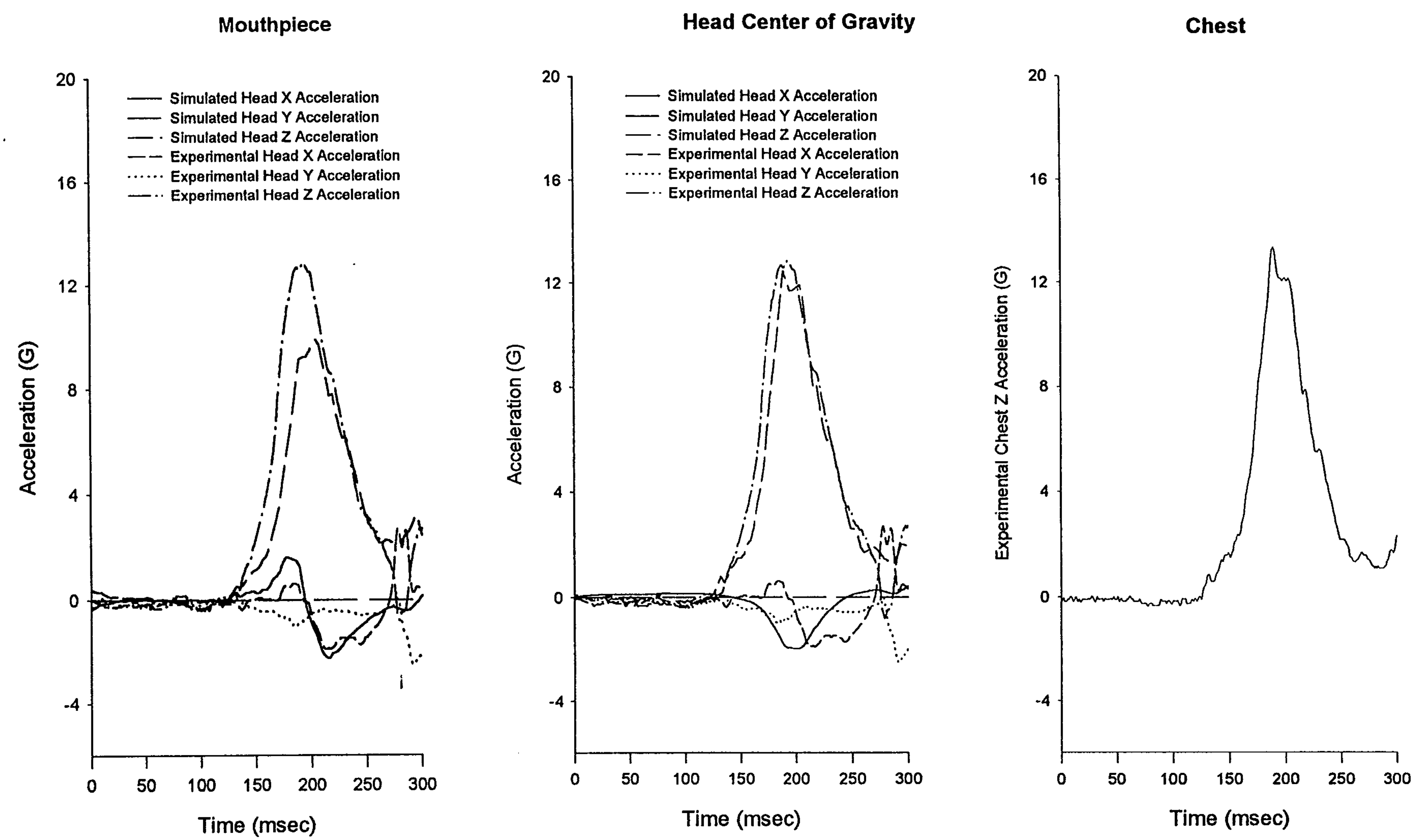


Figure 9 Acceleration data for mouthpiece, head center of gravity and chest for male subject 2309.

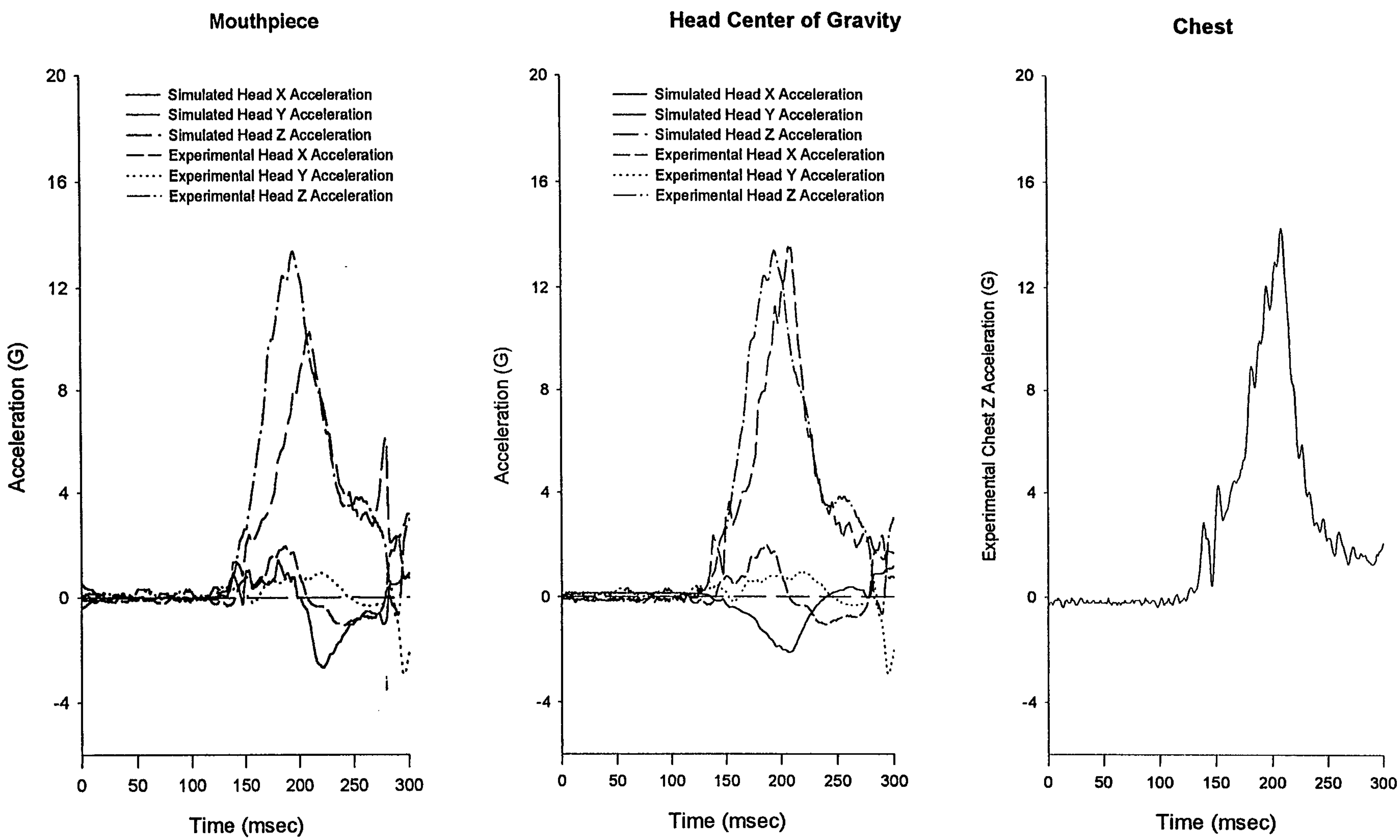


Figure 10 Acceleration data for mouthpiece, head center of gravity and chest for male subject 2429.

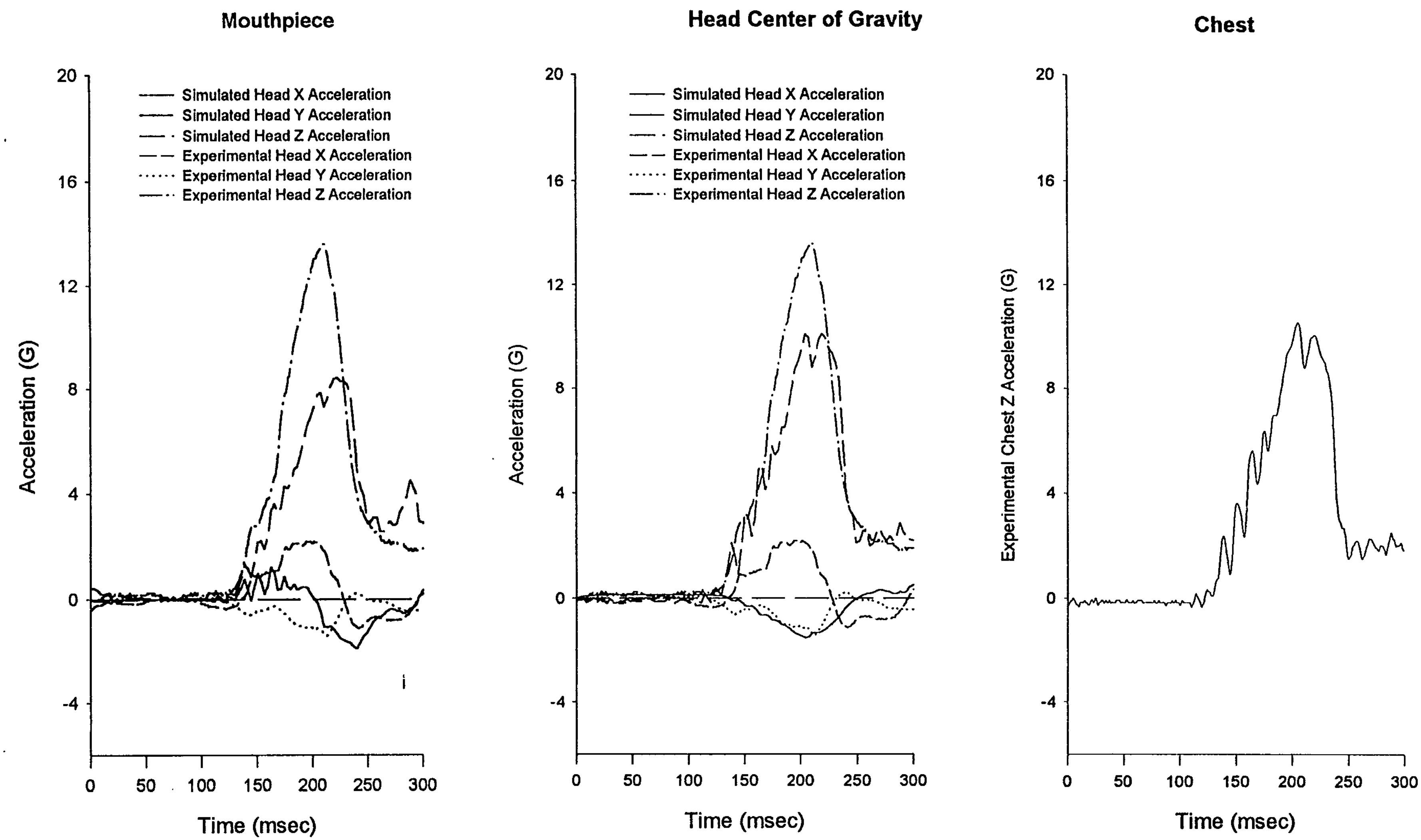


Figure 11 Acceleration data for mouthpiece, head center of gravity and chest for male subject 2442.

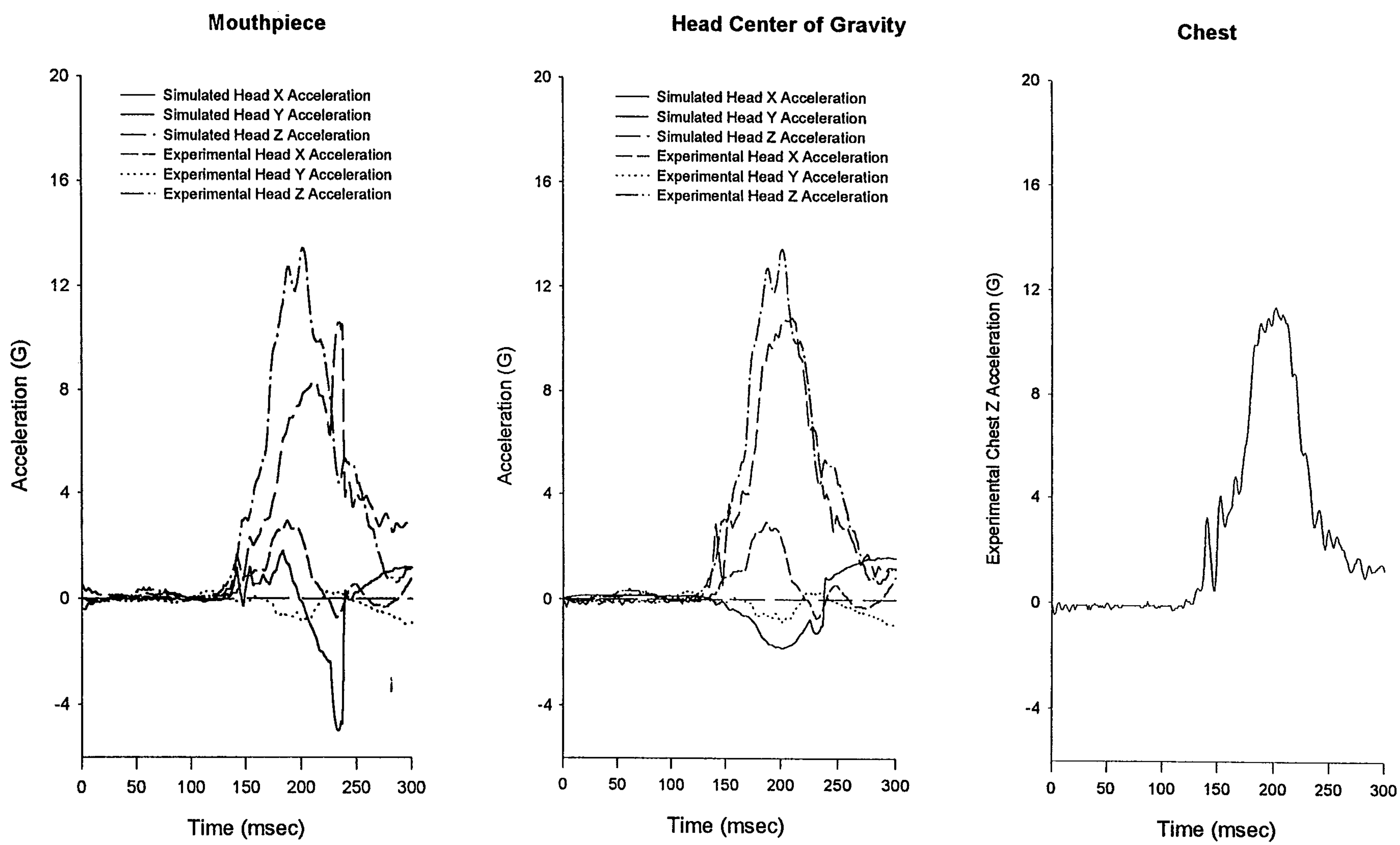


Figure 12 Acceleration data for mouthpiece, head center of gravity and chest for male subject 2504.